From Conformal Invariants to Percolation

C. McMullen

30 April, 2021

Contents

1 Introduction

The flexibility of conformal mappings is a cornerstone of geometry and analysis in the complex plane. Conformal invariance also underpins the remarkable properties of scaling limits of discrete random processes in two dimensions. This course will provide an introduction to foundational results and current developments in the field.

I. We will begin with a review of the flexibility and rigidity of conformal mappings, centered around the Riemann mapping theorem. This topic will also serve as a review of complex analysis. We will then move on to other conformal invariant such as the Poincaré metric on a plane region, extremal length, harmonic measure and capacity. Highlighting the central importance of harmonic functions, we will also review the solution to the Dirichlet problem.

Main references: [Ah1], [Ah2].

II. We will then turn to a discussion of random walks on lattices and discrete harmonic functions. We will then discuss the continuum limit of these discretized process, showing that (a) discrete harmonic functions limit to smooth ones; and (b) random walks limit to a natural continuous process, *Brownian motion*. We will also show that Brownian motion is conformally invariant, and that it leads to a useful, dynamic perspective on potential theory.

This discussion illuminates the remarkable fact that the continuum limit of a suitable random process on a lattice with rather few symmetries, can acquire many more in the continuum limit: it becomes conformally invariant.

Main references: [DY].

III. As a final illustration of this idea, we will discuss critical site percolation on the triangular lattice, and Smirnov's proof of the conformal invariance of the probability of crossings.

Main references: [BR], [Sm].

2 Conformal mappings and metrics

Standard domains. The complex plane, the unit disk, the upper halfplane and the Riemann sphere will be denoted by:

$$\begin{aligned}
\mathbb{C} &= \{z = x + iy \ : \ x, y \in \mathbb{R}\}; \\
\Delta &= \{z \in \mathbb{C} \ : \ |z| < 1\}; \\
\mathbb{H} &= \{z \in \mathbb{C} \ : \ \mathrm{Im}(z) > 0\}; \quad \text{and} \\
\widehat{\mathbb{C}} &= \mathbb{C} \cup \{\infty\}.
\end{aligned}$$

Let U be an open subset of $\mathbb{C}$. A function $f : U \to \mathbb{C}$ is *analytic* if $f'(z)$ exists at each point $z \in U$. When $f'(z) = a \neq 0$, the derivative of f acts on tangent vectors by multiplication by a. Thus Df scales lengths by $|a|$ and *preserves angles*.

Remarkably, the condition that $f'(z)$ exists implies that all derivatives of f exist, and that locally we can write f as a convergent power series: near any point $p \in U$ we have

$$f(z) = \sum_0^\infty a_n (z - p)^n.$$

Automorphisms and metrics. The automorphisms of $\widehat{\mathbb{C}}$ are given by Möbius transformations. Every automorphism of a standard domain extends to one of $\widehat{\mathbb{C}}$. Up to $\pm I$ we have $\mathrm{Aut}(\widehat{\mathbb{C}}) = \mathrm{SL}_2(\mathbb{C})$, $\mathrm{Aut}(\mathbb{H}) = \mathrm{SL}_2(\mathbb{R})$, and $\mathrm{Aut}(\Delta) = \mathrm{SU}(1,1)$.

A *conformal metric* on a domain U is given by $\rho(z)|dz|$ where $\rho(z) > 0$. One can integrate ρ along a smooth arc, and ρ^2 over a region, to obtain their length and area respectively.

Each standard region carries a natural metric: $|dz|$ and $2|dz|/(1 + |z|^2)$ on $\mathbb{C}$ and $\widehat{\mathbb{C}}$, and $2|dz|/(1 - |z|^2)$ and $|dz|/\mathrm{Im}(z)$ on $\mathbb{H}$. The latter two are invariant under $\mathrm{Aut}(\Delta)$ and $\mathrm{Aut}(\mathbb{H})$. Up to scale, they are the only invariant metrics.

The automorphisms of $\widehat{\mathbb{C}}$ preserving the spherical metric are $\mathrm{SU}(2)$.

Note: metrics pull back under analytic maps, by

$$f^* \rho = \rho(f(z))|f'(z)|\,|dz|.$$

The map $f(z) = i(1 - z)/(1 + z)$ gives an isomorphism $f : \Delta \to \mathbb{H}$ and one can check that $f^*(\rho_\mathbb{H}) = \rho_\Delta$.

The Schwarz Lemma. A powerful tool in complex analysis is provided by the fact that holomorphic maps tend to be *contractions.*

Theorem 2.1 *Let $f : (\Delta, 0) \to (\Delta, 0)$ be analytic. Then $|f(z)| \leq |z|$ for all $z \in \Delta$, and $|f'(z)| \leq 1$. If equality holds in either case, then $f(z) = e^{i\theta}z$ is a rotation.*

Proof. It is easy to see there is a unique analytic function $g : \Delta \to \overline{\Delta}$ satisfying $g(z) = f(z)/z$ for $z \neq 0$ (and $g(0) = f'(0)$). For $|z| \leq r < 1$ we have, by the maximum principle, $|g(z)| \leq 1/r$; taking the limit as $r \to 1$ we find $|g(z)| \leq 1$ for all z, and $g(z)$ is constant $e^{i\theta}$ when equality holds. The desired conclusion for $f(z)$ follows. ∎

Corollary 2.2 *Any holomorphic map of the disk to itself is either an isometry or a contraction for the hyperbolic metric.*

Corollary 2.3 *Any holomorphic map between hyperbolic Riemann surfaces is either a locally isometric covering map, or a contraction.*

Corollary 2.4 *A bounded entire function is constant.*

Remark: the Hilbert metric. Any convex region U in $\mathbb{R}^n$ carries a (possibly degenerate) *Hilbert metric*, defined by

$$d(p, q) = \log[x, p, q, y]$$

where x and y are the endpoints of the line $\overline{xy} \cap U$, and $[x, p, q, y]$ is the cross–ratio, normalized so that $[0, 1, \lambda, \infty] = \lambda$. An affine (or even real projective) map from one convex region to another does not expand the Hilbert metric; often it contracts.

For the region bounded by an ellipse in $\mathbb{R}^2$, the Hilbert metric is isomorphic to the hyperbolic metric.

Interpolation. Consider the following problem: given (a_i, b_i), $i = 1, \ldots, n$, can we find an analytic map $f : \Delta \to \Delta$ such that $f(a_i) = b_i$?

For $n = 1$ the answer is always yes: we can take f to be a Möbius transformation.

For $n = 2$ the answer is yes, *provided* the Schwarz Lemma is satisfied: $d(a_1, a_2) \geq d(b_1, b_2)$. When equality holds, f is unique, otherwise it is flexible.
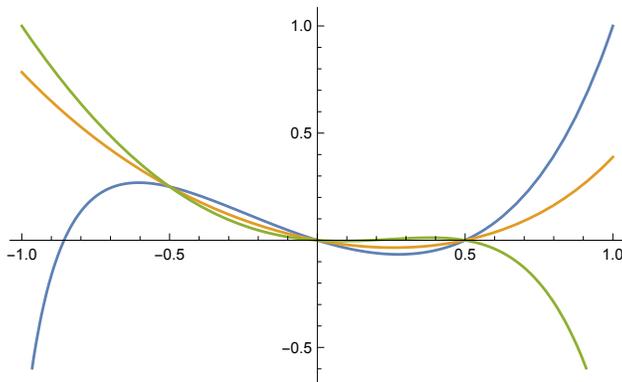
Figure 1. Interpolation: $f : \Delta \to \Delta$ with $f(-1/2) = 1/10$ and $f(0) = f(1/2) = 0$.

See Figure 1 for a flexible example when $n = 3$. A succinct criterion for interpolation was given by Pick. Here we will simply describe how the problem can be solved recursively. An important role will be played by the automorphisms of the disk of the form

$$M_a(z) = (z - a)/(1 - \bar{a}z),$$

$a \in \Delta$. Note that the inverse of M_a is M_{-a}.

To see the usefulness of these maps, note that the *Blaschke product*

$$B(z) = \prod_1^n M_{a_i}(z)$$

solves the interpolation problem $f(a_i) = 0$ for $i = 1, 2, \ldots, n$. The general solution to this interpolation problem is given by $f(z) = g(z)B(z)$ for any function $g : \Delta \to \overline{\Delta}$. Note that the constant functions with values on S^1 are included.

Let $\mathcal{F}$ denote the space of all analytic $g : \Delta \to \overline{\Delta}$, and let $\mathcal{F}(a, b)$ denote those satisfying $f(a_i) = b_i$.

Suppose $f \in \mathcal{F}(a, b)$. Then $M_{b_n}(f(z))$ has a zero at a_n. Thus we can divide by $M_{a_n}(z)$ to get a new function

$$g(z) = M_{b_n}(f(z))/M_{a_n}(z). \tag{2.1}$$

6

By considering the boundary values, we see that $g \in \mathcal{F}$. Moreover, g solves a new interpolation problem: for $i = 1, \ldots, n - 1$, setting $a_i' = a_i$, we have

$$g(a_i') = M_{b_n}(b_i)/M_{a_n}(a_i) = b_i'.$$

We can also reverse equation (2.1) above to solve for f, given g. Thus the general solution is given by

$$f(z) = M_{-b_n}(g(z)M_{a_n}(z)),$$

where $g \in \mathcal{F}(a', b')$.

Note that it might happen that $|b_i'| > 1$: in this case the original interpolation problem has no solution. It might also happen that $|b_i'| = 1$ for some i; in this case, the only possible solution is a constant function, and in particular we must also have $b_i = b_j$ for $j = 1, \ldots, n - 1$.

Conformal maps. A *conformal map* is an analytic homeomorphism $f : U \to V$. When U or V is Δ or $\mathbb{H}$, we often refer to f as a Riemann mapping.

For concreteness, consider a Riemann mapping $f : U \to \mathbb{H}$. Such a map is unique up to post-composition with $g \in \operatorname{Aut}(\mathbb{H})$. The Poincaré metric on U is given by

$$\rho_U = f^*(\rho_{\mathbb{H}}) = \frac{|f'(z)|\,|dz|}{\operatorname{Im} f(z)},$$

since $\rho_{\mathbb{H}} = |dz|/\operatorname{Im}(z)$. More generally, if we have a conformal *covering map* $\pi : \mathbb{H} \to V$, then the deck group $\Gamma \subset \operatorname{Aut}(\mathbb{H})$ acts isometrically, and there is a unique metric on V such that π is a local isometry from $\rho_{\mathbb{H}}$ to ρ_V.

Riemann mappings are very flexible, as we will see below: any region $U \subset \mathbb{C}$ that is *homeomorphic* to Δ is conformally isomorphic to Δ – except for $U = \mathbb{C}$. This coexistence of rigidity and flexibility is the hallmark of analysis in one complex variable.

Examples. To illustrate this flexibility, let us examine some concrete examples.

1. Möbius transformations. Any region $U \subset \mathbb{C}$ bounded by a circle admits a conformal map $f : U \to \Delta$ given by a Möbius transformation. The same is true for any halfplane. This allows its hyperbolic metric to be easily computed; e.g. on $\Delta(R) = \{z : |z| < R\}$ we have

$$\rho_{\Delta(R)} = \frac{2R|dz|}{R^2 - |z|^2}.$$

7

The region

$$1/\Delta = \{z \in \widehat{\mathbb{C}} \; : \; |z| > 1\}$$

is isomorphic to Δ via the map $z \mapsto 1/z$, which gives

$$\rho_{1/\Delta} = \frac{2|dz|}{|z|^2 - 1}.$$

2. The important mapping $f(z) = z + 1/z$ sends $1/\Delta$ to $\widehat{\mathbb{C}} - [-2, 2]$. As a variant, for $0 < A < 1$, $f(z) = z + A/z$ sends $1/\Delta$ to the *exterior* of an ellipse.

3. Let $S = \{z \; : \; 0 < \mathrm{Im}(z) < \pi\}$. We then have a conformal isomorphism

$$f : \mathbb{H} \to S$$

given simply by $f(z) = \log z$. Here we have taken the standard branch of the logarithm, with $\arg(z) \in (0, \pi)$. The inverse map is given by exponentiation. Using the fact that for $z = x + iy$, we have

$$|e^z| = e^x \quad \text{and} \quad \mathrm{Im}(e^z) = \mathrm{Im}(e^x(\cos y + i \sin y)) = e^x \sin y,$$

we find

$$\rho_S = \frac{|dz|}{\sin(y)}.$$

4. The map $f(z) = z^\alpha$, $0 < \alpha \le 2$, sends $\mathbb{H}$ to the region

$$H_\alpha = \{z \; : \; 0 < \arg(z) < \pi\alpha\}.$$

To make $f(z)$ well–defined, we require that it extends the usual map $x \mapsto x^\alpha$ on $\mathbb{R}$.

5. A *lune L* is the lemon-shaped region between a pair of circles. A Möbius transformation can be chosen to send the tips of the lune to 0 and ∞; then after a suitable rotation, L becomes the region H_α above, and it can then be mapped to $\mathbb{H}$.

In particular, the lune $\Delta \cap \mathbb{H}$ is isomorphic to $\mathbb{H}$.

6. Let $\Delta^* = \Delta - \{0\}$. Then $f(z) = \exp(iz)$ provides a *covering map*

$$f : \mathbb{H} \to \Delta^*$$

with deck group $\mathbb{Z}$ generated by $z \mapsto z + 2\pi$. The inverse map is given locally by $g(z) = (1/i)\log(z) = i\log(1/z)$, and thus $|g'(z)| = 1/|z|$ and $\text{Im}\, g(z) = \log(1/|z|)$. This gives

$$\rho_{\Delta^*} = \frac{|g'(z)|\,|dz|}{\text{Im}\, g(z)} = \frac{|dz|}{|z \log |z||}.$$

One should compare this metric to the *cylindrical metric* $|dz|/|z|$, which gives infinite volume near the puncture at $z = 0$. The metric above has a finite volume end. Note also that $p : z \mapsto z^2$ is a covering map of Δ^* to itself, and hence a local isometry. Thus the circle $|z| = r^2$ has half the length of the circle $|z| = r$, in the hyperbolic metric on Δ^*.

By iterating p on the region $r^2 \le |z| \le r$, we obtain a sequence of annuli $A_0, A_1, A_2, \ldots$ filling up a neighborhood of the cusp at $z = 0$ and with $\text{area}(A_i) = 2^{-i}\,\text{area}(A_0)$. This gives another way to see that $\bigcup A_i = \{z : 0 < |z| < r\}$ has finite area.

7. Let S_+ be the half–infinite rectangle

$$S_+ = (0, \infty) + i(0, \pi).$$

Note that $\exp(S_+)$ is the region in $\mathbb{H}$ with $|z| > 1$. Thus we can apply the map $f(z) = (z + 1/z)/2$ to $\exp(S_+)$ to send it to $\mathbb{H}$. The upshot is that

$$\cosh : S_+ \to \mathbb{H}$$

is a conformal isomorphism.

8. By similar reasoning, for $S' = (-\pi/2, \pi/2) + i(0, \infty)$, the sine function gives a conformal map

$$\sin : S' \to \mathbb{H}.$$

9. Now let us study the inverse map: we have

$$f(z) = \sin^{-1}(z) = \int \frac{dz}{\sqrt{1 - z^2}}.$$

9

The derivative of this map is real on $[-1, 1]$ but imaginary on the rest of $\partial\mathbb{H}$. The square–root in the denominator introduces a *bend* in the map.

More generally, the *Schwarz–Christoffel formula* says that for $a_1 < a_2 < \cdots < a_n$ and $\beta_i \in (0, 1)$, the function

$$f(z) = \int \frac{dz}{\prod_1^n (z - a_i)^{\beta_i}}$$

gives a conformal map from $\mathbb{H}$ to a convex region with external angles $\pi\beta_i$. For example,

$$f(z) = \int \frac{dz}{\sqrt{z(1 - z)(1 + z)}}$$

sends $\mathbb{H}$ to the interior of a square. What is the side length of this square? It is given by the definite integral

$$L = \int_0^1 (x - x^3)^{-1/2} \, dx = \frac{2\sqrt{\pi}\Gamma(5/4)}{\Gamma(3/4)}.$$

The heuristic here is that the map $f(z) = z^\alpha$ transforms the $\mathbb{H}$ to a region with an internal angle of $\pi\alpha$ at the origin, and an external angle $\beta = \pi(1 - \alpha)$; and $f'(z) = \alpha z^{\alpha-1}$ is proportional to $1/z^\beta$.

10. It interesting to consider the mapping $f(z) = z^{i\alpha}$ with $\alpha > 0$. On the positive real axis we have

$$f(x) = \exp(i\alpha \log x) = \exp(it)$$

where $t = \alpha \log x$ runs from $-\infty$ to ∞; thus f sends $\mathbb{R}_+$ to S^1 by a covering map. Similarly on the negative real axis, we have

$$f(-x) = \exp(i\alpha(\log x + \pi i)) = \exp(-\pi\alpha)f(x).$$

Thus $-\mathbb{R}_+$ is sent to the circle of radius

$$r = \exp(-\pi\alpha),$$

also by a covering map. In fact, f sends $\mathbb{H}$ to the annulus $A(r)$ defined by $r < |z| < 1$ by a covering map, and this presentation of the universal cover can be used to compute the hyperbolic metric on an the annulus. By considering the length of its core geodesic, one can show:

$A(r)$ *is isomorphic to* $A(s)$ *iff* $r = s$.

The length is given by $L = \log \lambda$, for the smallest $\lambda > 1$ such that $f(\lambda z) = f(z)$; explicitly, $L = 2\pi/\alpha$.

The Riemann mapping theorem. The prime example of flexibility in conformal geometry in $\mathbb{C}$ is:

Theorem 2.5 *Let* $U \subset \mathbb{C}$ *be a simply–connected region,* $U \neq \mathbb{C}$. *Then for each* $p \in U$, *there exists a unique conformal isomorphism*

$$F : (U, p) \to (\Delta, 0)$$

such that $F'(p) > 0$.

Proof of the Riemann mapping theorem (sketch). Consider

$$\mathcal{F} = \{f : (U, p) \to (\Delta, 0) : f \text{ is an injective analytic function with } f'(p) > 0\}.$$

Then there is a unique $F \in \mathcal{F}$ that maximizes $f'(p)$, and this map is an isomorphism to its image. ∎

It is a useful exercise to fill in all the steps in the proof.

1. First, why is $\mathcal{F}$ nonempty? For example, what if

$$(U, p) = (\mathbb{C} - [1, \infty), 0)?$$

For this we use the fact that whenever $\pi_1(U)$ is trivial and $0 \notin U$, there exists an analytic function

$$\log : U \to \mathbb{C}$$

such that $\exp(\log(z)) = z$. Given $a \in U$, this function can be defined by

$$\log(u) = \log(a) + \int_a^u \frac{dz}{z},$$

where the integral is along any path in U from a to U. The fact that U is simply–connected guarantees that the integral does not depend on the choice of path.

We can then define $z^\alpha = \exp(\alpha \log z)$ for any complex α, and in particular there is a single–valued branch of $\sqrt{z}$ defined on U.

Since $U \neq \mathbb{C}$, we can first translate U so $0 \notin U$; then replace U with its image $V = \sqrt{U}$ under a single–valued branch of $\sqrt{z}$. Then $V \cap -V = \emptyset$. Pick $p \in -V$; then $f(z) = 1/(z - p)$ sends V to a bounded set, and rescaling we get an open subset of the unit disk.

2. Next, we would like to show that there exists an $f \in \mathcal{F}$ that maximizes $f'(p)$. To see this, we first recall that Cauchy's integral formula gives, for function analytic in $B(z, r)$:

$$f'(z) = \frac{1}{2\pi i} \int_{\partial B(z,r)} \frac{f(\zeta)\,d\zeta}{(z - \zeta)^2}.$$

Since the length of $\partial B(z, r)$ is $2\pi r$, this gives

$$|f'(z)| \leq \frac{\max_{B(z,r)} |f|}{r}.$$

Since $|f| \leq 1$, this shows $|f'(z)| \leq 1/d(z, \partial U)$ for all $z \in U$.

Thus by the Arzela–Ascoli theorem, any sequence $f_n \in \mathcal{F}$ has a subsequence converging uniformly on compact sets. The limit $g = \lim f_n$ is analytic by Cauchy's theorem.

3. We claim that if g is not constant, then $g \in \mathcal{F}$. The key point is that g is univalent. This can be proved by the argument principle: if $f|B = B(z, r)$ is nowhere vanishing, then the number of zeros of f in B is given by

$$N(f) = \frac{1}{2\pi i} \int_{\partial B} \frac{f'(z)\,dz}{f(z)}.$$

Suppose $g(z_1) = z_2$, where $z_1 \neq z_2$ and both lie in U. Since g is nonconstant, there is a small ball $B = B(z_1, r)$ excluding z_2 such that $g(z) - g(z_2)$ has no zeros on ∂B. Then $N(g(z) - g(z_2)) \geq 1$. But $N(g(z) - g(z_2)) = \lim N(f_n(z) - f_n(z_2))$, so for n large enough there is a $z_1' \in B$ such that $f_n(z') = f_n(z_2)$, contradicting injectivity of f_n. Thus g is injective.

4. Let $r = d(p, \partial U) > 0$. By the Schwarz Lemma, $f'(p) \leq 1/r$ for all $f \in \mathcal{F}$. Let $f_n'(p) \to \sup_{\mathcal{F}} f'(p)$. By the argument above, after passing to a subsequence $g = \lim f_n \in \mathcal{F}$, so this supremum is achieved.

To complete the proof, we must show that $g(U) = \Delta$.

5. If not, we can choose an $A \in \mathrm{Aut}(\Delta)$ such that $0 \notin V = A(g(U))$. Let $s(z) = z^2$. Since V is simply–connected, we can also choose a single–valued branch of s^{-1} on V. Finally we can choose $B \in \mathrm{Aut}(\Delta)$ such that
$$f = B \circ s^{-1} \circ A \circ g : U \to \Delta$$
satisfies $f(0) = 0$ and $f'(0) > 0$. Then $f \in \mathcal{F}$.

Now $g = M \circ f$, where
$$M = A^{-1} \circ s \circ B^{-1} : (\Delta, 0) \to (\Delta, 0).$$

By the Schwarz Lemma, $|M'(0)| < 1$. Thus $g'(p) < f'(p)$, contradicting the definition of g.

Thus g is a bijection to the unit disk. By the Schwarz Lemma, it is unique.

Application to dynamics. Using the Schwarz Lemma and the Riemann mapping theorem, we have:

Theorem 2.6 *Let $f : \mathbb{C} \to \mathbb{C}$ be a polynomial map of degree $d > 1$. Then every attracting periodic cycle of f attracts a critical point.*

Corollary 2.7 *The number of periodic attracting cycles is at most $d - 1$.*

For details see e.g. [Mil].

Boundary values. A *Jordan curve* is a continuum $J \subset \mathbb{C}$ homeomorphic to S^1. A *Jordan domain* is a bounded, connected open set $U \subset \mathbb{C}$ with ∂U a Jordan curve.

Theorem 2.8 *Let $F : \Delta \to U \subset \mathbb{C}$ be a conformal map to a bounded domain. Then the radial limit*
$$\lim_{r \to 1} F(rz)$$
exists for almost every $z \in S^1$.

Proof. We have $\int_\Delta |F'|^2 < \infty$; by Cauchy–Schwarz, the length of the image of almost every ray of Δ under F is finite, and thus radial limits exist. ∎

Corollary 2.9 *If U is a Jordan domain, then F extends to a homeomorphism between $\overline{\Delta}$ and $\overline{U}$ sending S^1 to ∂U.*

Proof. Apply the fact that small area implies all arcs have small length to a neighborhood of a point in S^1, to construct a small crosscut. This gives continuity of F. If $F(z_1) = F(z_2)$ for $z_1 \neq z_2$ in S^1, we get that F is constant along an arc $[z_1, z_2]$, a contradiction. Thus F is injective and hence a homeomorphism. ∎

Corollary 2.10 *If ∂U is locally connected, then F extends continuously to S^1.*

Wild domains. It is clear that we cannot expect the Riemann mapping $f : U \to \Delta$ to extend continuously to ∂U in general; consider, for example, the case where $U = \Delta - [0, 1]$. We also cannot expect boundary extension for $F : \Delta \to U$ in general; for this to exist, ∂U must be locally connected. The results above are thus best possible.

Topology in $\mathbb{R}^2$ and $\mathbb{R}^3$. The Riemann mapping theorem gives an analytic proof of a topological theorem, namely that any simply–connected open region $U \subset \mathbb{R}^2$ is homeomorphic to a disk. This result is *false* in $\mathbb{R}^3$.

It is more usual to work in S^3. Whitehead produced the following example of a *contractible open set* $U \subset S^3$ such that U is not homeomorphic to a ball:

$$U = \bigcup U_i = \bigcup_1^\infty S^3 - T_i,$$

where $T_1 \supset T_2 \supset T_3 \ldots$ is a nested sequence of solid tori such that:

1. Each T_i is unknotted in S^3; and hence $\pi_1(U_i) \cong \mathbb{Z}$;

2. The torus T_{i+1} is knotted in T_i; but

3. The meridian of T_i has zero linking number with T_{i+1}.

Then $\pi_2(U_i) = 0$ for all i and $\pi_1(U_i)$ maps to zero in $\pi_1(U_{i+1})$, so $\pi_1(U) = \pi_2(U) = 0$ and thus U is contractible by basic homotopy theorem. But a loop L linking the Whitehead continuum $W = \bigcap T_i$ does not bound an *embedded* disk $D \subset U$, so U is not homeomorphic to a ball.

For more details see e.g. [Rol].

The hyperbolic metric and the $1/d$ metric. (Cf. [Neh, V.8].)

Let $U \subset \mathbb{C}$ be a simply–connected domain, $U \neq \mathbb{C}$. An easily–understood metric on U is the $1/d$ metric, defined by

$$\rho_{1/d}(z) = \frac{1}{d(z, \partial U)}.$$

Note that the hyperbolic metric $\rho_{\mathbb{H}}$ is exactly equal to the $1/d$ metric on $\mathbb{H}$, and in all the examples we have examined, ρ_U blows up like $1/d$ near ∂U. In fact this is a general phenomenon:

Theorem 2.11 *For any simply–connected domain $U \subset \mathbb{C}$, $U \neq \mathbb{C}$, and any $z \in U$, we have*

$$\rho_U(z)/\rho_{1/d}(z) \in [1/2, 2].$$

The bound in one direction is immediate from the Schwarz Lemma: let $z \in U$, $r = d(z, \partial U)$ and $B = B(z, r) \subset U$. Since this inclusion is a contraction from the hyperbolic metric, we have

$$\rho_B(z) = 2\rho_{1/d}(z) \geq \rho_U(z).$$

Spaces of univalent maps. For the proof in the other direction, it is useful to consider the space of all Riemann mappings, with some mild normalizations. Two spaces are commonly considered:

The space S is the space of injective analytic maps $f : (\Delta, 0) \rightarrow (\mathbb{C}, 0)$, normalized so that $f'(z) = 1$. Thus an element of f has the form

$$f(z) = \sum_1^\infty a_n z^n = z + a_2 z^2 + \dots$$

Each f is a Riemann mapping to its image, $U = f(\Delta)$. We give S the topology of uniform convergence on compact sets.

The space Σ consists of injective analytic maps from the outside of the closed unit disk to the outside of a full compact set $K = \widehat{\mathbb{C}} - f(1/\Delta)$. (Here *full* means $\mathbb{C} - K$ is connected). These maps

$$f : (\mathbb{C} - \overline{\Delta}) \rightarrow (\mathbb{C} - K(f))$$

are normalized so that

$$f(z) = z + \frac{b_1}{z} + \frac{b_2}{z^2} + \cdots = z + \sum_1^\infty b_n z^{-n}.$$

Again, the topology is uniform convergence on compact sets.

The letters S and Σ come from the German word *schlicht* (simple), which means univalent in this context.

Here are some basic properties of these spaces:

1. The spaces S and Σ are *compact.*

2. (The area theorem.) For $f \in \Sigma$, we have

$$\mathrm{area}(K) = \pi(1 - \sum n|b_n|^2).$$

   In particular we have
$$\sum n|b_n|^2 \le 1.$$

3. (The Bieberbach conjecture/de Brange theorem.) For $f \in S$ we have $|a_n| \le n$.

4. (The Koebe 1/4 theorem.) For any $f \in S$, we have

$$f(\Delta) \supset B(0, 1/4).$$

(In fact, $d(0, \partial f(\Delta)) \in [1/4, 1]$.)

Since $f'(0) = 1$, for $U = f(\Delta)$ we have

$$\rho_U(0) = \rho_\Delta(0) = 2.$$

Thus the final statement implies Theorem 2.11.

Extremal maps. An important map in the space S is the one that achieves equality for both the Bieberbach conjecture and the Koebe 1/4 theorem; it is given by
$$f(z) = z + 2z^2 + 3z^3 + \cdots = \frac{z}{(1-z)^2}.$$

(Note that $(1 - z)^{-2}$ is the derivative of $1/(1 - z) = 1 + z + z^2 + \cdots$.) Here $f(\Delta)$ is the complement of the arc $(-\infty, -1/4]$.

16

In Σ an important role is played by

$$f(z) = z + 1/z;$$

its image is the complement of $K = [-2, 2]$.

The area theorem and its consequences. Let us sketch the proofs. Suppose $K \subset \mathbb{C}$ is a smoothly bounded disk. We then have

$$\text{area}(K) = \frac{i}{2} \int_K dz \wedge d\bar{z} = \frac{1}{2i} \int_{\partial K} \bar{z} \wedge dz.$$

Suppose $f \in \Sigma$, f is smooth and injective on S^1 and $\partial K = f(S^1)$. Using the (orientation–preserving) parameterization $f : S^1 \to \partial K$, we have

$$\text{area}(K) = \frac{1}{2i} \int_{S^1} \bar{f}(z) \, df,$$

and the result follows using the fact that $\bar{z} = 1/z$ on S^1 and $\int_{S^1} z^n \, dz = 2\pi i$ if $n = -1$ and 0 otherwise.

From the area theorem we obtain $|b_n|^2 \leq 1/n$ which yields compactness of Σ. The area theorem also yields the first instance of the Bieberbach conjecture:

Theorem 2.12 *For $f \in S$ we have $|a_2| \leq 2$.*

Proof. First we compute that if $f \in S$ then

$$1/f(1/z) = z(1 + a_2 z^{-1} + a_3 z^{-2} + \cdots)^{-1}$$
$$z - a_2 + (a_2^2 - a_3) z^{-1} + \cdot.$$

(Here we have used $(1+w)^{-1} = 1 - w + w^2 - w^3 + \cdots$). Thus $1/f(z) + a_2 \in \Sigma$, so $|a_2^2 - a_3| \leq 1$. Next we observe that $\sqrt{f(z^2)}$ is also in S, and satisfies

$$\sqrt{f(z^2)} = z(1 + a_2 z^2 + a_3^4 + \cdots)^{1/2}$$
$$= z + (a_2/2) z^3 + \ldots.$$

(Here we have used $(1+w)^{1/2} = 1 + (1/2)w - (1/8)w^2 + \cdots$). Hence $|a_2| \leq 2$.
∎

The Koebe 1/4 theorem follows. If $f \in S$ and $p \notin f(\Delta)$, then $A(f(z)) \in S$ too, where $A(z) = z/(1 - z/p)$; and we have

$$A(f(z)) = (z + a_2 z^2 + \cdots)(1 + z/p + \cdots) = z + (a_2 + 1/p)z^2 + \cdots,$$

which gives

$$2 \geq |a_2 + 1/p| \geq |1/p| - 2$$

and hence $|1/p| \leq 4$.

Since a_2 is bounded, we can use the map $f \mapsto 1/f(1/z) + a_2 : S \to \Sigma$ to derive compactness of S from compactness of Σ. Compactness yields immediately qualitative bounds, uniform over S, such as $|a_n| = O(1)$ and $\sup_{|z|<r} |f^{(n)}(z)| \leq M(n, r)$.

For example, $f(\Delta)$ can be a wild domain, but the image of $S^1(r)$, $r < 1$, is a smooth, analytic loop with curvature bounded by $k(r)$ independent of f. (The white of the egg may splatter, but the yolk is always good.)

The triply–punctured sphere. Let $U = \mathbb{C} - \{0, 1\}$. It is well–known from the theory of elliptic functions that the universal cover of U is naturally isomorphic to the upper halfplane.

In fact there is a covering natural map

$$\lambda : \mathbb{H} \to \mathbb{C}$$

whose image is U. To define this map, one associates to every $\tau \in \mathbb{H}$ a complex torus

$$E = \mathbb{C}/\mathbb{Z} \oplus \mathbb{Z}\tau,$$

and a basis for the points of order 2 on E, namely $e_1 = 1/2$ and $e_2 = \tau/2$. Setting $e_0 = 0$ and $e_3 = e_1 + e_2$, we have $E[2] = \{e_0, e_1, e_2, e_3\}$.

Next the Weierstrass $\wp$ function gives a natural degree two map

$$\wp : E \to \widehat{\mathbb{C}}.$$

Up to composition with an element of $\operatorname{Aut} \widehat{\mathbb{C}}$, $\wp$ is unique determined by the fact that $\wp(z) = \wp(-z)$. Finally, λ is defined as the cross–ratio:

$$\lambda(\tau) = [\wp(e_0), \wp(e_1), \wp(e_2), \wp(e_3)].$$

The value of λ is canonically attached to the complex torus E and an ordering of $E[2]$. Thus it is not quite invariant under $\operatorname{SL}_2(\mathbb{Z})$; it is invariant under

$$\Gamma(2) = \{A \in \operatorname{SL}_2(\mathbb{Z}) \ : \ A = I \bmod 2\}.$$

It establishes an isomorphism

$$\lambda : \Gamma(2)\backslash\mathbb{H} \cong \mathbb{C} - \{0,1\}.$$

Consequently, we have a natural hyperbolic metric ρ_U on $U = \mathbb{C} - \{0,1\}$. The fact that the universal cover of U is the disk easily leads to the small and great Picard theorems:

Theorem 2.13 *An entire function $f : \mathbb{C} \to \mathbb{C}$ omits at most one value.*
An analytic function omits at most one value in $\mathbb{C}$ in any neighborhood of an essential singularity.

3 Harmonic functions and harmonic measure

Harmonic functions are close cousins of analytic functions. In this section we recall their basic properties, and discuss the role of conformal mappings in solving the Dirichlet problem. The solution produces a family of conformally natural *harmonic measures* on the boundary of any region.

The Laplacian on $\mathbb{R}^n$. Let us denote the inner product on $L^2(\mathbb{R}^n)$ by

$$\langle \phi_1, \phi_2 \rangle = \int_{\mathbb{R}^n} \phi_1(x)\phi_2(x)\, dx,$$

where $dx = dx_1 \ldots dx_n$ is the volume form, and ϕ_1 and ϕ_2 are real–valued functions.

The *Laplacian* of a C^2 function on $\mathbb{R}^n$ is given by

$$\Delta\phi = \sum \frac{d^2\phi}{dx_i^2}.$$

It is closely related to the *Dirichlet energy*, defined for compactly supported smooth functions by:

$$D(\phi) = \int_{\mathbb{R}^n} |\nabla\phi|^2\, dx.$$

Integrating by parts, we find

$$D(\phi) = -\langle \phi, \Delta\phi \rangle.$$

Thus $-\Delta$ is the self–adjoint operator that relates the Dirichlet energy to the L^2–norm.

The function ϕ is *harmonic* if $\Delta\phi = 0$.

A harmonic function ϕ on a domain Ω formally minimizes its Dirichlet energy, given its boundary conditions. That is, if we perturb ϕ by a compactly supported function ψ, its energy remains constant to first order: we have

$$\frac{d}{ds} D(\phi + s\psi) = \int \nabla\psi \cdot \nabla\phi = -\int \psi(\Delta\phi) = 0.$$

Flows and charge. The Laplacian can also be written as

$$\Delta\phi = \nabla \cdot \nabla\phi.$$

The function ϕ can be interpreted as the *potential* for an electrical charge distribution given by the *measure*

$$\mu = (\Delta\phi)\, dx.$$

The associated electrical field is $\nabla\phi$. This vector field describes the force on a unit charge, due to the charge distribution μ.[1]

By Stoke's theorem, the total charge contained in a compact region $\overline{\Omega}$ is the same as the *flux* of the electric field through its boundary; we have:

$$\text{flux}(\nabla\phi, \partial\Omega) = \int_{\partial\Omega} \nabla\phi \cdot \widehat{n}\, dS = \int_\Omega \Delta\phi = \mu(\Omega).$$

The maximum principle. In particular, the gradient of a harmonic function generates a *volume preserving* flow; its flux is zero. On the other hand, it is easy to see that

$$\frac{d}{dr} \text{avg}_{S^{n-1}(p,r)}\, \phi = C(r)\, \text{flux}(\nabla\phi, S^{n-1}(p,r)),$$

and thus the average of a harmonic of a harmonic function over a sphere centered at p is the same as $\phi(p)$. The same is true for the average over balls, since this is a weighted sum of averages over spheres. Summing up, we have:

Theorem 3.1 (Mean value property) *Let $\phi(x)$ be harmonic in a domain $\Omega \subset \mathbb{R}^n$. Then for any ball $B = B^n(p,r)$ contained in Ω, we have*

$$\phi(p) = \text{avg}_B\, \phi(x) = \frac{1}{|B|} \int_B \phi(x)\, dx.$$

The same is true with B replaced by the sphere of radius r about p.

---

[1] In physics the force is usually given by $-\nabla\phi$, so a particle moves to decrease its potential energy; but then one also writes $\Delta\phi = -\mu$.

Corollary 3.2 (Maximum Principle) *A harmonic function on a connected region does not achieve its maximum unless it is constant.*

Proof. Let $M = \sup_\Omega \phi(x)$, and let $F \subset \Omega$ be the closed set where $\phi(x) = M$. By the maximum principle, F is also open: if $p \in M$, then M is also the average of ϕ over a small ball B about p; since $\phi|B$ is bounded by M, we must have $B \subset F$. Since Ω is connected, either F is empty (and the maximum is not achieved), or $F = \Omega$ (and ϕ is constant). ∎

The same is true for the minimum.

Corollary 3.3 *If a harmonic function ϕ is continuous up to the boundary of a compact region $\overline{\Omega}$, then*

$$\max_{\partial\Omega} |\phi| = \max_{\overline{\Omega}} |\phi|.$$

Physical intuition. Here are some statements from physics that are useful tests of one's intuition. We use the fact that the gravitational potential also satisfies Laplace's equation.

1. There is no way to hold an electron stably at a point with a static field.

2. The gravitational field outside the earth is the same as if all its mass were concentrated at the center.

3. The gravitational field vanishes inside a massive spherical shell.

(The first and last follow from the maximum principle; the second, by spherical symmetry and consideration of flux.)

Weyl's Lemma. Let ψ_r be a C^∞, rotationally invariant function on $\mathbb{R}^n$, supported on $|x| < r$, with $\int \psi = 1$. Let ϕ be a C^2 harmonic function on Ω.

By the mean–value property,

$$\phi * \psi_r(x) = \phi(x)$$

provided $B(x, r) \subset \Omega$. But this convolution is C^∞; thus ϕ itself is C^∞.

By similar reasoning, one can show that any distribution satisfying $\Delta\phi = 0$ is actually a smooth function. This is a manifestation of the ellipticity of the Laplace operator, called *Weyl's Lemma*.

Limits; converse mean value property. Because of the reasoning above, any continuous or even locally integrable function $\phi : \Omega \to \mathbb{R}$ that satisfies the mean–value property is actually *smooth*. In fact we have:

Theorem 3.4 *A locally integrable function on Ω with the mean value property for balls $B \subset \Omega$ is harmonic.*

Proof. It suffices to prove $\Delta\phi = 0$ at $x = 0$, assuming this point is in Ω. By Taylor's theorem we can write $\phi(x) = \phi_0(x) + \phi_1(x) + \phi_2(x) + \epsilon(x)$, where $\phi_d(x)$ is homogeneous of degree d. The mean value property must hold for each term. Now $\phi_2(x) = \sum a_{ij} x_i x_j$. For $i \neq j$, $x_i x_j$ is odd under $x_i \to -x_i$, so its average over a sphere $|x| = r$ is zero. Thus the mean value property implies that the average of $\sum a_{ii} x_i^2$ over the sphere is zero. This average is a constant multiple of $2 \sum a_{ii} = (\Delta\phi)(0)$. ∎

Since the mean value varies continuously under uniform limits, we have:

Corollary 3.5 *A uniform limit of harmonic functions is harmonic, and under uniform convergence all derivatives converge as well.*

Proof. If $\phi_n \to \phi$ uniformly, then we also have

$$D^\alpha \phi_n = \phi_n * (D^\alpha \psi_r) \to D^\alpha \phi$$

uniformly. ∎

Compactness for harmonic functions. Since a harmonic function can be locally expressed as $\phi = \phi * \psi_r$, the sup-norm of ϕ locally controls all of its derivatives. This shows:

Theorem 3.6 *The set of harmonic functions with $\sup |\phi| \leq 1$ on a region $\Omega \subset \mathbb{R}^n$ is compact in the topology of uniform convergence on compact sets.*

Subharmonic functions. A function is *subharmonic* if $\Delta\phi \geq 0$. Such functions satisfy the mean–value inequality:

$$\phi(p) \leq \mathrm{avg}_{S^{n-1}(p,r)}\, \phi(x),$$

by the same reasoning as above; and hence they satisfy the maximum principle (but not the minimum principle). If $\Delta\phi \leq 0$, we say ϕ is *superharmonic*.

When $n = 1$, harmonic functions are linear, and subharmonic functions are *convex*. Convexity provides a useful intuition for their behavior.

It is conventional to allow subharmonic functions to assume the value $-\infty$. In particular, if $f(z)$ is analytic, then $\phi(z) = \log|f(z)|$ is subharmonic. (If f has no zeros, then ϕ is harmonic.)

The Laplacian in complex analysis. In the complex plane with coordinates $z = x + iy$, we have

$$\Delta\phi = \frac{d^2\phi}{dx^2} + \frac{d^2\phi}{dy^2}.$$

Alternatively, the Laplacian of a function can be described naturally as a 2-form, given by

$$(\Delta\phi)\, dx \wedge dy = d * d\phi = (2i)\, \partial\bar{\partial}\phi.$$

This version of the Laplacian is invariant under analytic change of coordinates; it is the right notion on a Riemann surface.

We will explain each of these expressions in turn, and use them to establish the following basic connection between harmonic and analytic functions:

Theorem 3.7 *In the complex plane, a real–valued function $\phi(z)$ is harmonic if and only if $\phi(z)$ is locally the real part of an analytic function $f(z)$.*

The Hodge star. The complex structure on a domain $\Omega \subset \mathbb{C}$ gives rise to an algebraic operator on 1–forms, defined by

$$*dx = dy, \quad *dy = -dx.$$

Note that $*^2 = -1$. The Hodge star on 1–forms is conformally natural.

In polar coordinates, we have

$$*dr = r\, d\theta, \quad *d\theta = dr/r.$$

The Laplacian of a function is then *naturally* the 2–form defined by

$$d * d\phi = d(\phi_x\, dy - \phi_y\, dx) = (\phi_{xx} + \phi_{yy})\, dx \wedge dy = (\Delta\phi)\, dx \wedge dy.$$

The *flux* of $\nabla\phi$ through the (oriented) boundary of a region is given naturally by

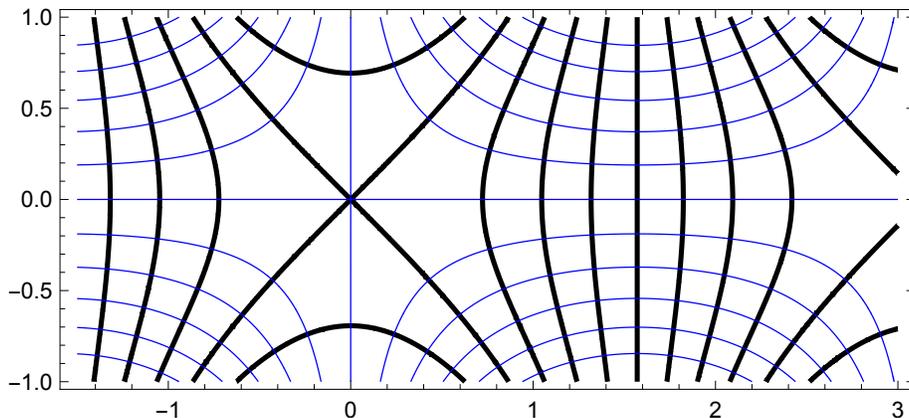$$\text{flux}(d\phi, \partial\Omega) = \int_{\partial\Omega} *d\phi.$$

Figure 2. Orthogonal foliations for $\cos(z)$.

Foliations and harmonic conjugation. A harmonic function gives rise naturally to a *pair of orthogonal foliations.* These foliations come from the closed 1–forms $d\phi$ and $*d\phi$. More precisely, the first foliation is given by the level sets of ϕ; for the second, we use the fact that $d*d\phi = 0$ to *locally* construct a function ψ such that

$$d\psi = *d\phi; \tag{3.1}$$

then *its* level sets form the second foliation. Equation (3.1) ensures these foliations are orthogonal. The function ψ is called the *harmonic conjugate* for ϕ; it is well–defined up to an additive constant.

One can check that $f(z) = \phi(z) + i\psi(z)$ is a analytic. The foliations just described are the preimage under f of the foliations of $\mathbb{C}$ by horizontal and vertical lines. For an example, see Figure 2.

The ∂ and $\overline{\partial}$ operators. For a different perspective on the connection between analytic and harmonic functions, it is useful to introduce the differential operators

$$\frac{df}{dz} = \frac{1}{2}\left(\frac{df}{dx} - i\frac{df}{dy}\right) \quad \text{and} \quad \frac{df}{d\overline{z}} = \frac{1}{2}\left(\frac{df}{dx} + i\frac{df}{dy}\right).$$

When applied to polynomials in $\mathbb{C}[x, y] = \mathbb{C}[z, \overline{z}]$, they behave as if z and $\overline{z}$ are independent variables. The Cauchy–Riemann equations give:

f *is analytic* $\iff$ $df/d\overline{z} = 0$;

24

and the Laplacian satisfies

$$\Delta f = 4 \frac{d^2 f}{dz\, d\bar{z}}.$$

More intrinsically, any 1–form on a region in $\mathbb{C}$ can be naturally expressed as a sum

$$\alpha = \alpha_1\, dz + \alpha_2\, d\bar{z};$$

these are the $(1,0)$ and $(0,1)$ parts of α. Similarly, for any smooth function ϕ we can write:

$$d\phi = \partial\phi + \overline{\partial}\phi = (d\phi/dz)\, dz + (d\phi/d\bar{z})\, d\bar{z}.$$

The operator $d = \partial + \overline{\partial}$ also acts on 1–forms, and satisfies $d^2 = \partial^2 = \overline{\partial}^2$. We find that:

$$\partial\overline{\partial}\phi = \frac{d^2 f}{dz\, d\bar{z}}\, dz \wedge d\bar{z}$$

is naturally a 2–form. Since

$$dx \wedge dy = \frac{i}{2}\, dz \wedge d\bar{z},$$

this gives

$$(\Delta\phi)\, dx \wedge dy = 2i\, \partial\overline{\partial}\phi.$$

Proof of Theorem 3.7. If f is analytic, then $\partial\overline{\partial}f = 0$, and similarly for $\overline{f}$; thus the real and imaginary parts of f are harmonic.

Conversely, if ϕ is harmonic, then $\omega = \partial\phi$ is a closed form, since $\overline{\partial}\omega = 0$. Thus locally $\omega = df$, and $\overline{\partial}f = 0$, so f is analytic. Finally

$$d\phi = \omega + \overline{\omega} = d(f + \overline{f}),$$

so there exists a constant such that $\phi = 2\,\mathrm{Re}(f) + c.$ ∎

Conformal invariance of Dirichlet energy. The *Dirichlet energy* of a complex–valued function f on a domain Ω is given by:

$$D(f) = \int_\Omega |\nabla f|^2\, |dz|^2 = \int df \wedge *d\overline{f}$$

The final expression is conformally natural.

Thus one can canonically attach to any Riemann surface X the inner product space of functions with finite Dirichlet energy, and its Hilbert space

completion. (One should take the quotient by the constant functions, since these have norm zero.)

Note that the usual L^2-norm of a function,

$$\int |f(z)|^2 \, |dz|^2,$$

is *not* a conformal invariant; it implicitly uses the Euclidean metric. The first formula for $D(\phi)$ *mixes* the Euclidean metric, and the gradient; it is a remarkable fact that this mixture is invariant.

Here is an interpretation of the Dirichlet energy, making its invariance less mysterious.

Theorem 3.8 *Let $f(z) = \alpha(z) + i\beta(z)$ be an analytic function on Ω. Then*

$$D(\alpha) = D(\beta) = (1/2)D(f) = \mathrm{area}(f(\Omega)),$$

where the area of the image is computed with multiplicity.

Proof. Note that $dz \wedge *d\bar{z} = 2dx \wedge dy$ is twice the standard area form on $\mathbb{C}$. Since $D(f)$ is just the integral of the pullback of this form, we have $D(f) = 2\,\mathrm{area}(f(\Omega))$. The other equalities follow by a simple calculation. ∎

This inner product space is the starting point for the construction of the Gaussian free field, to be discussed in a later section.

Harmonic functions and metrics. In the presence of a conformal metric $\rho = \rho(z)\,|dz|$, we also have a Hodge star operator from 2–forms to functions, characterized by

$$*\rho(z)^2 \, dx \wedge dy = 1.$$

We then obtain a Laplacian sending functions to functions, defined by

$$\Delta_\rho \phi = *d * d\phi.$$

In local coordinates this gives

$$\Delta_\rho \phi(z) = \rho^{-2}(z)\,\Delta\phi(z).$$

For the Euclidean metric $\rho(z) = 1$, we recover the usual Laplacian.

Fundamental solution. A *fundamental solution* to Laplace's equation is provided by

$$E(z) = \log|z|.$$

This function has the property that

$$\Delta E = 2\pi\delta_0,$$

as can be seen by considering its flux: we have $*dE = *(dr/r) = d\theta$, and $\int_{S^1(r)} d\theta = 2\pi$ for any $r > 0$.

The fundamental solution allows one to solve the equation $\Delta\phi = 2\pi g$; the solution is given by $\phi = E * g$. In fact for any measure μ, we have

$$\Delta(E * \mu) = 2\pi\mu.$$

Since the logarithmic singularity of μ is very mild, the potential $\phi = E * \mu$ is often continuous; for example, it is continuous if $\mu(B(x, r)) = O(r^\delta)$ for some $\delta > 0$.

When μ is a compactly probability supported measure, the associated potential satisfies

$$(E * \mu)(z) = \log|z| + O(1/|z|) \tag{3.2}$$

as $|z| \to \infty$, because

$$E(z + a) = \log|z + a| = \log|z| + \log|1 + a/z| = E(z) + O(1/|z|).$$

For a general measure with compact support, we have

$$(E * \mu)(z) = \mu(\mathbb{C})\log|z| + O(1/|z|).$$

The Dirichlet problem. Let $U \subset \widehat{\mathbb{C}}$ be an open region. The *Dirichlet problem* for U is the following: given a continuous function Φ on ∂U, extend ϕ continuously to $\overline{U}$ in such as way that $\phi|U$ is harmonic. By the maximum principle, the solution ϕ is unique if it exist, and we have

$$\|\phi\|_\infty \le \|\Phi\|_\infty.$$

When the Dirichlet problem has a solution for all Φ, the map $P(\Phi) = \phi$ defines a bounded linear operator

$$P : C(\partial U) \to C(\overline{U}).$$

27

Heuristically, ϕ is the unique function that minimizes the Dirichlet energy

$$\int_U |\nabla \phi|^2$$

among all extensions of Φ. However in practice, the solution to the Dirichlet problem may have infinite energy.

Solution on the unit disk. The main obstruction to solving the Dirichlet problem is that parts of ∂U may be too thin. For example, the Dirichlet problem has no solution on Δ^*, because any bounded harmonic function on Δ^* extends to one on Δ. The value of $\Phi(0)$ must equal the average of Φ over S^1 for the extension to exist.

However we do have:

Theorem 3.9 *The Dirichlet problem has a solution on the unit disk Δ.*

Proof. Let $F \subset C(S^1)$ be the space of finite Laurent polynomials,

$$\Phi(z) = \sum_{-N}^{N} a_n z^n = a_0 + \sum_1^N a_n z^n + \sum_1^N a_{-n} \overline{z}^n.$$

The expression at the left is harmonic in Δ; take is as the definition of $P(\Phi)$. Since F is dense in $C(S^1)$, and uniform limits of harmonic functions are harmonic, E has a unique continuous extension to $C(S^1)$ and it solves the Dirichlet problem. $\blacksquare$

The Poisson kernel. Note that by the mean value property, we know the value of $\phi(0)$ explicitly: it is given by

$$\phi(0) = \frac{1}{2\pi} \int_{S^1} f(z) |dz|.$$

By naturality, we also have for any $A \in \mathrm{Aut}(\Delta)$,

$$P(\Phi(Az)) = (P(\Phi))(Az) = \phi(Az).$$

This allows one to write down a formula for $\phi(z)$ at any point. This formula is called the *Poisson kernel.* Before writing down the formula, it is useful to explain its meaning geometrically:

Theorem 3.10 *The value of $\phi(z)$ is the visual average of $\Phi|S^1$ as seen from $z \in \Delta$.*

In other words, $\phi(z)$ is the expected value of $\Phi(\zeta)$ at the endpoint ζ of a randomly chosen hyperbolic ray based at z.

The ray at z is specified by its *initial tangent vector* at z, which is chosen randomly in S^1.

Proof. This result is simply the mean value property, restated so that is invariant under $\mathrm{Aut}(\Delta)$. ∎
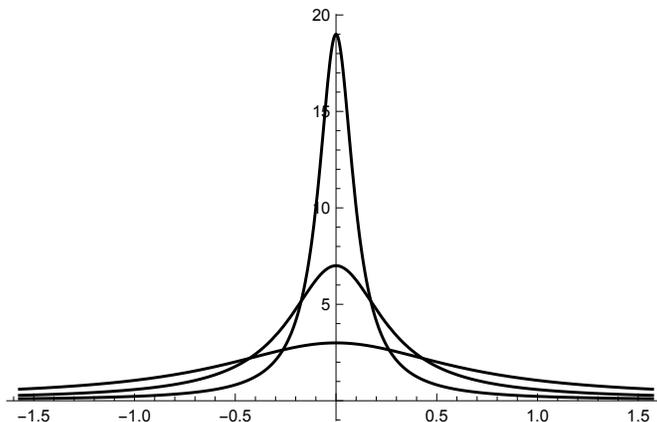

Figure 3. The Poisson kernel $P_r(\theta)$ for $r = 1/2, 3/4, 9/10$.

To write down the Poisson kernel explicitly, suppose Φ is not a continuous function but that the distribution δ_1 concentrated at $z = 1 \in S^1$. Then we expect ϕ to be a harmonic function such that $\phi(0) = 1$ (the total mass of Φ), and $\phi = 0$ on all of S^1 except for $z = 1$. Moreover, we have $\phi = \mathrm{Im}\, f(z)$ for a unique analytic function with $f(0) = i$.

The natural choice for f is the unique Möbius transformation sending $(\Delta, 0)$ to $(\mathbb{H}, i)$ with $f(1) = \infty$, namely

$$f(z) = i(1 + z)/(1 - z).$$

This choice gives

$$P(z) = \mathrm{Im}\, f(z) = \frac{1 - |z|^2}{|1 - z|^2}.$$

We then have:

29

Theorem 3.11 *The harmonic extension of Φ in $C(S^1)$ is given by*

$$\phi(z) = \frac{1}{2\pi} \int_0^{2\pi} \Phi(\zeta)P(z/\zeta)\,|d\zeta|.$$

The Poisson kernel is often written as a function of $z = re^{i\theta}$; it then becomes:

$$P_r(\theta) = \frac{1 - r^2}{1 - 2r\cos\theta + r^2}.$$

See Figure 3. Since the harmonic extension of the constant function $\Phi = 1$ is $\phi = 1$, we have $(1/2\pi)\int P_r(\theta)\,d\theta = 1$ for every r. The functions P_r are smooth approximations to the δ-function, and we have

$$\phi(re^{i\theta}) = (P_r * \Phi)(\theta) = \frac{1}{2\pi} \int_0^{2\pi} \Phi(t)P_r(\theta - t)\,dt.$$

Corollary 3.12 *The endpoint of a random hyperbolic geodesic in the unit disk, based on $0 < r < 1$, is distributed on S^1 according to the measure $(1/2\pi)P_r(\theta)\,|d\theta|$.*

We also note that formally, for $z \in S^1$ the series representation of the delta function is given by:

$$\delta_1(z) = \sum_{-\infty}^{\infty} z^n,$$

and $P(z)$ is obtained from the expression above by replacing the negative powers of z with $\bar{z}$:

$$
\begin{aligned}
P(z) &= \sum_0^{\infty} z^n + \sum_1^{\infty} \bar{z}^n = \frac{1}{1-z} + \frac{1}{1-\bar{z}} - 1 \\
&= \frac{2 - z - \bar{z}}{|1-z|^2} - 1 = \frac{1 - |z|^2}{|1-z|^2}.
\end{aligned}
$$

Extending measurable functions and distributions. Since $P_r(\theta)$ is smooth for each $r < 1$, it can be paired with an arbitrary distribution the circle. In this way every distribution has a natural harmonic extension to the disk. In particular, L^∞ functions on S^1 extend to bounded harmonic functions on the disk. This extension is continuous in the weak* topology; since $C(S^1)$ is weakly dense in $L^\infty(S^1)$, it is the unique such extension.

In fact, letting $H^\infty(\Delta) \subset C(\Delta)$ denote the closed subspace of all bounded harmonic functions, we have:

Theorem 3.13 *Convolution with the Poisson kernel gives an isometric isomorphism*

$$P : L^\infty(S^1) \to H^\infty(\Delta).$$

This extension is characterized by the property that for *almost every $z \in S^1$*, we have

$$\lim_{r \to 1-} \phi(rz) = \Phi(z).$$

Example. Let $\Phi : \widehat{\mathbb{R}} = \partial\mathbb{H} \to \mathbb{C}$ be given by $\Phi(x) = 0$ on $\mathbb{R}_+$ and $\Phi(x) = 1$ on $\mathbb{R}_-$. Its harmonic extension is given by:

$$\phi(z) = \arg(z)/\pi.$$

Solution on a topological disk. The Riemann mapping for a region U, when known explicitly, allows one to solve the Dirichlet problem there as well, as can be seen by the proof of:

Theorem 3.14 *The Dirichlet problem has a solution on any Jordan domain $U \subset \mathbb{C}$.*

Proof. Let $F : \overline{\Delta} \to \overline{U}$ be the extended Riemann mapping. Then F gives an isomorphism between $C(\overline{U})$ and $C(\overline{\Delta})$ preserving harmonic functions, so the Poisson extension operator P_Δ pulls back to give a solution to the Dirichlet problem on U. Explicitly,

$$P_U(\Phi) = (P_\Delta(\Phi \circ F)) \circ F^{-1}.$$

$\blacksquare$

Example: Fluid flow. Since idealized fluid flow preserves volume, conformal mapping can also be used to describe streamlines in a simply–connected region. An example is show in Figure 4. The region $U = \mathbb{H} - B$, where B is a ball of radius 1 resting on the real axis. The streamlines are the images of horizontal lines in $\mathbb{H}$ under a suitably normalized map $f : \mathbb{H} \to U$.

Piecewise–continuous boundary data. Here is a generalization of the maximum principle that will be useful when studying harmonic functions whose boundary values are allowed to have jumps at finitely many points.

Note that we add the *a priori* assumption that ϕ is bounded. The Poisson kernel $P(z)$ shows that some assumption of this kind is required: its boundary values are 0 except for one point on S^1.
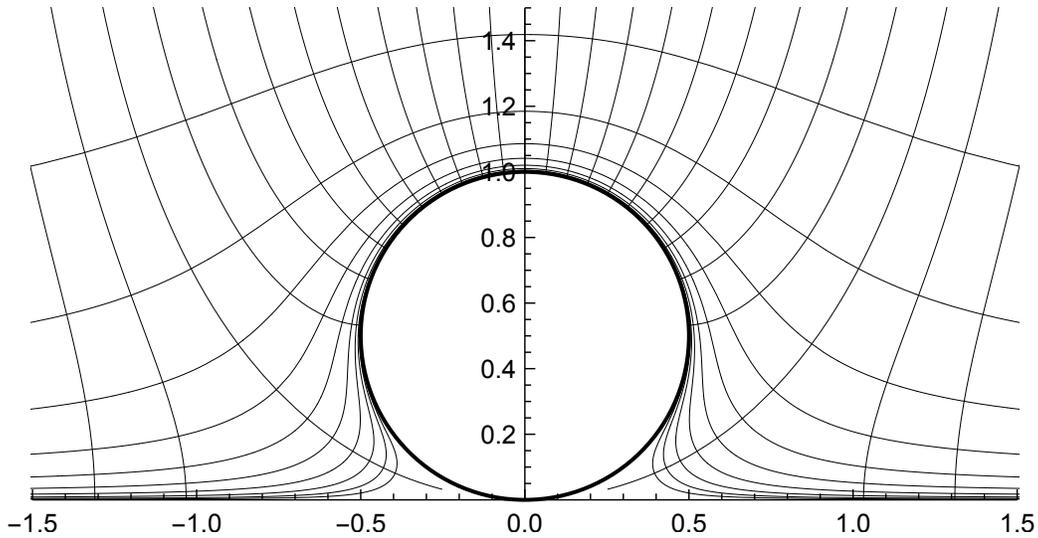
Figure 4. Streamlines around an obstacle.

Theorem 3.15 (Lindelöf) *Suppose ϕ is bounded and harmonic on $\Omega \subset \mathbb{C}$, with boundary values $\leq M$ outside a finite set $\{a_1, \ldots, a_n\} \subset \partial\Omega$. Then*

$$\sup_{\Omega} \phi(z) \leq M.$$

Proof. Assume Ω is bounded; the general case is handled by a limiting argument. Let $R = \operatorname{diam}(\Omega)$ and consider, for any $s > 0$, the modified harmonic function

$$\phi_s(z) = \phi(z) + s \sum_{1}^{n} \log(|z - a_i|/R) \leq \phi(z).$$

Since ϕ is bounded, the function ϕ_s tends to $-\infty$ at the points a_i, so by the maximum principle it is bounded by M in Ω. The theorem follows by letting $s \to 0$. ∎

Since the Dirichlet problem on Δ can be solved for bounded measurable functions on S^1, one expects a similar result for other simply connected domains. As a step in this direction, we note:

Theorem 3.16 *The Dirichlet problem has a unique* bounded *solution for piecewise–continuous boundary data on any Jordan domain U.*
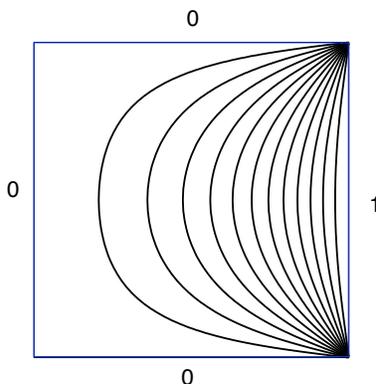
32

Figure 5. Harmonic function on a square.

The boundary data is continuous apart from jump at finitely many points; it is automatically bounded. Uniqueness comes from Lindelöf's theorem.

Example: temperature distribution. Figure 5 shows a harmonic function $\phi(z)$ on the unit square with boundary values 1 on one edge, and 0 on the remaining edges. This function describes the distribution of temperature in a refrigerated compartment with the door open.

This function is given by

$$\phi(z) = u(\wp(\omega z; g_2, g_3))$$

where $\wp(z; g_2, g_3)$ is the Weierstrass $\wp$ function for the cubic equation $4t^3 - g_2 t - g_3 = 4(t^3 - t)$, invariant under the lattice $\Lambda = \omega(\mathbb{Z} \oplus i\mathbb{Z})$; and $u(z) = \pi^{-1}(\arg(z) - \arg(z - 1))$.

Solution on a finitely–connected region. The Dirichlet problem for a region $\Omega \subset \mathbb{C}$ has a robust solution in many other cases.

A typical case where it does *not* have a solution is when $\Omega = \mathbb{C} - K$ and K is a very thin Cantor set. In this case, every bounded harmonic function on Ω is constant, so there can be no extension.

As a useful intermediate case, we will show:

Theorem 3.17 *The Dirichlet problem has a bounded solution for piecewise–continuous functions on the boundary of a region $\Omega \subset \mathbb{C}$ bounded by finitely many Jordan curves.*

33

By Lindelöf's maximum principle, the boundedness condition makes the solution unique.

The alternating method. The proof of Theorem 3.17 will use *Schwarz's alternating method*, which leverages the solution on a Jordan domain. This approach emphasizes our clear understanding for the unit disk, and the Riemann mapping theorem. (Another approach would be to use *Perron's method* and subharmonic functions; this method generalizes to $\mathbb{R}^n$.)

Proof. To begin, choose a finite set of arcs $A \subset \Omega$ running between points on $\partial\Omega$ such that $U = \Omega - A$ is simply–connected. Let $V = \Omega - B$ for a disjoint set of similar arcs B. We can then solve the Dirichlet problem for suitable functions on

$$\partial U = A \cup \partial\Omega \quad \text{and} \quad \partial V = B \cup \partial\Omega.$$

Now let $\Phi : \partial\Omega \to \mathbb{R}$ be a piecewise–continuous function. Since Φ is bounded, we can normalize so that $0 \le \Phi \le 1$.

We will construct a sequence of functions u_i and v_i that solve the Dirichlet problem on U and V respectively. For all i we set $u_i|\partial\Omega = v_i|\partial\Omega = \Phi$. However the boundary data on A and B will vary with i.

To begin, let u_1 solve the Dirichlet problem on U with $u_1|A = 1$. We can regard u_1 as a continuous function on Ω that is harmonic outside A. Then, let v_1 solve the Dirichlet problem on V with $v_1|B = u_1|B$. Continue inductively, so that u_i and v_i are harmonic on U and V respectively, continuous on Ω, and

$$u_i|A = v_{i-1}, \quad v_i|B = u_i.$$

We claim that

$$1 \ge u_1 \ge v_1 \ge u_2 \ge v_2 \ge \cdots \ge 0$$

throughout Ω. The proof is by induction. By the maximum principle, it suffices to check these inequalities on $A \cup B$.

The first case is easy: we have $v_1|B = u_1|B$ and $v_1 \le 1 = u_1|A$ by the maximum principle, so $v_1 \le u_1$.

Now we address u_2. By assumption, $u_2 = v_1 \le u_1$ on A. This implies $u_2 \le u_1$ everywhere in Ω. In particular,

$$u_2|B \le u_1|B = v_1|B.$$

Thus $u_2 \le v_1$. The same argument handles the general inductive step.

Because these function are decreasing, we can define:

$$\phi(z) = \lim u_i(z) = \lim v_i(z).$$

By compactness of bounded harmonic functions, the limit is harmonic on $\Omega - (A \cup B)$. But the u_i are also harmonic outside B, and the v_i are harmonic outside A, so ϕ is harmonic on Ω.

It remains to show that if f is continuous on a neighborhood of $p \in \partial\Omega$, then ϕ is a continuous extension of f. For this we choose a neighborhood U of p such that $U \cap \Omega$ is a Jordan domain, and use the Riemann mapping theorem to reduce to the case of $\phi = \lim u_n$ on a disk. ∎

To visualize the proof above, it is useful to keep in mind the case where $\Omega = \{r < |z| < 1\}$ is an annulus, $A = [r, 1]$ and $B = [-1, -r]$.

Harmonic measure. Let $\Omega \subset \widehat{\mathbb{C}}$ be a region for which we have a solution to the Dirichlet problem,
$$P : C(\partial\Omega) \to \Omega.$$

Then every $z \in \Omega$ determines a bounded linear functional on $C(\partial\Omega)$, by evaluating the harmonic extension $\phi = P(\Phi)$ at z. Boundedness follows from the maximum principle.

By the Riesz representation theorem, there is a unique Borel probability measure ω_z on $\partial\Omega$ characterized by

$$\phi(z) = \int_{\partial\Omega} \Phi(s)\,\omega_z.$$

The measure of a Borel set $A \subset \partial\Omega$ is often denoted

$$\omega_z(A) = \omega(z, \Omega, A).$$

We will denote the measure itself by $\omega(z, \Omega)$. When A is a finite union of arcs, $\omega(z, \Omega, A)$ is the unique harmonic function on Ω that tends to 1 on A and to 0 elsewhere on ∂A, the endpoints of A being excluded.

Examples.

1. Harmonic measure from the origin in the unit disk is given by normalized angular measure on the circle:

$$\omega(0, \Delta) = |d\theta|/2\pi.$$

35

Thus
$$\omega(0, \Delta, A) = |A|/2\pi.$$

As we have seen, harmonic measure from a general point $z \in \Delta$ is the hitting measure for a random hyperbolic geodesic ray based at z. It is given by the Poisson kernel: for $r > 0$,
$$\omega(r, \Delta) = (1/2\pi)P(re^{i\theta})\,|d\theta|.$$

2. For $z \in \mathbb{H}$ and $A = [a, b] \subset \mathbb{R}$, we have
$$\omega(z, \mathbb{H}, A) = \alpha/\pi,$$

where α is the angle at z of the triangle with vertices (a, z, b).

Indeed, a harmonic function equal to 1 on $A = [a, b] \subset \mathbb{R}$ and zero elsewhere is given by
$$\omega(z, \mathbb{H}, [a, b]) = (1/\pi)(\arg(z - b) - \arg(z - a)).$$

Note that α is $(1/2)$ of the *visual angle* β of A in the hyperbolic metric, which comes from the *hyperbolic triangle* with vertices (a, z, b).

3. We have
$$\omega(i, \mathbb{H}) = \frac{1}{\pi}\frac{|dx|}{1 + x^2}.$$

(The integral of this density is $(1/\pi)\tan^{-1}(x)$.) Thus the Poisson operator on $\mathbb{H}$ satisfies
$$\phi(i) = \frac{1}{\pi}\int_{\mathbb{R}}\frac{\Phi(x)\,dx}{1 + x^2},$$

and a similar formula can be given by $\phi(z)$ at any $z \in \mathbb{H}$.

4. Let $\Omega = A(a, b) = \{z : a < |z| < b\}$. Rather than computing harmonic measure on $\partial A(a, b)$, we just consider the measure of its inner and outer boundary components A and B, where $|z| = a$ and $|z| = b$ respectively. First we note that:
$$\omega(z, A(a, b), B) = \frac{\log(r/a)}{\log(b/a)},$$

where $|z| = r$. Thus if $r = a^{1-p}b^p$, and we set $s = b/a$, the fraction above becomes $\log(s^p)/\log s$ and we obtain
$$\omega(a^{1-p}b^p, A(a, b), B) = p.$$

36

Monotonicity. It is useful to extend the definition of harmonic measure to allow A to be any closed subset of $\overline{\Omega}$. In this case for $z \in \Omega' = \Omega - A$ we define

$$\omega(z, \Omega, A) = \omega(z, \Omega', A \cap \partial\Omega'),$$

and $\omega(z, \Omega, A) = 1$ if $z \in A$. We then have, by the maximum principle, the following result:

Theorem 3.18 *The harmonic measure $\omega(z, \Omega, A)$ increases if A increases or if Ω increases.*

For example, let $\gamma \subset \mathbb{H}$ be a path connecting zero to infinity. Let Ω be the component of $\mathbb{H} - \gamma$ whose closure contains $\mathbb{R}_+$. Let $A \subset \mathbb{H}$ be the closed set bounded by γ and $\mathbb{R}_-$. We then have

$$\omega(z, \Omega, \gamma) = \omega(z, \mathbb{H}, A) \geq \omega(z, \mathbb{H}, \mathbb{R}_-) = \arg(z)/\pi.$$

Majorization. A principal application of harmonic measure is to controlling the size of harmonic and analytic functions in Ω in terms of their behavior on $\partial\Omega$. As a simple example we have:

Theorem 3.19 *Let $\phi(z)$ be a bounded harmonic function in Ω, let $\partial\Omega = A \sqcup B$, and suppose $\phi(z) \leq m$ on A and $\leq M$ on B. We then have*

$$\phi(z) \leq pm + qM,$$

where $p = \omega(z, \Omega, A)$ and $q = 1 - p = \omega(z, \Omega, B)$.

Proof. Apply the maximum principle, using the fact that

$$\phi(z) \leq m\omega(z, \Omega, A) + M\omega(z, \Omega, B)$$

on $\partial\Omega$. ∎

The log modulus function. For analytic functions a somewhat stronger conclusion holds. Let $f : \Omega \to \mathbb{C}$ be a nonzero analytic function. To analyze the behavior of $|f|$, we will make use of the function

$$\phi(z) = \log |f(z)| = \operatorname{Re} \log f(z).$$

This function is harmonic if f has no zeros. In general, it is subharmonic; in fact, we have:
$$\Delta \log|f| = \sum_{f(a)=0} 2\pi\delta_a,$$
where the zeros are summed with multiplicity. Indeed, if f has a zero of multiplicity n at a, then we can write $f(z) = (z-a)^n g(z)$ where $g(a) \neq 0$, and then
$$\log|f(z)| = n\log|z-a| + \log|g(z)|,$$
and the final term is harmonic near a.

Theorem 3.20 *Let $f(z)$ be a bounded analytic function on Ω and $\partial\Omega = A \sqcup B$. Suppose $|f(z)| \leq m$ in A and $|f(z) \leq M$ on B. We then have*
$$|f(z)| \leq m^p M^{1-p}$$
in the region where $\omega(z, \Omega, A) \geq p$.

Proof. Apply the principle above to the subharmonic function $\phi(z) = \log|f(z)|$. ∎

Corollary 3.21 *If an analytic function vanishes on a set of positive harmonic measure in $\partial\Omega$, then it vanishes identically.*

Hadamard's 3–circles theorem. We will use majorization as above to prove Hadamard's theorem. Let $f(z)$ be analytic in the annulus $r_1 < |z| < r_2$, and let
$$M(r) = \sup_{|z|=r} |f(z)|.$$

Theorem 3.22 *The function $\log M(r)$ is convex as a function of $\log r$.*

Proof. Consider a, b, c such that $r_1 < a < b < c < r_2$. Let Ω be the annulus bounded by the circles A and C where $|z| = A$ and $|z| = C$ respectively. Write $b = a^p c^{1-p}$, with $0 < p < 1$, and suppose $|z| = b$. Then $\omega(z, \Omega, A) = p$, and thus
$$|f(z)| \leq M(a)^p M(c)^{1-p}.$$
Taking the sup over $|z| = b$ and then logarithms, we find
$$\log M(b) \leq p\log M(a) + (1-p)\log M(c).$$
Since $\log b = p\log a + (1-p)\log c$, we have established convexity. ∎

Radial limits. Let $f : \mathbb{H} \to \mathbb{C}$ be an analytic function. We say f has a *radial limit c* at infinity if whenever $z_n \to \infty$ while staying within a bounded hyperbolic distance of the imaginary axis, we have $f(z_n) \to c$. This is equivalent to the condition that $\arg(z_n)$ is bounded away from 0 and π.

Clearly the radial limit is unique if it exists.

We say c is an *asymptotic value* of f if $f(z) \to c$ as $z \to \infty$ along some path $\gamma \subset \mathbb{H}$.

Theorem 3.23 *Suppose c is an asymptotic value of a bounded analytic function $f : \mathbb{H} \to \mathbb{C}$. Then c is also the radial limit of f at infinity.*

Corollary 3.24 *A bounded analytic function can have a most one asymptotic value.*

Corollary 3.25 *Let $f : \Delta \to U$ be a Riemann mapping, and let $\gamma \subset U$ be a path converging to a point $p \in \partial U$. Let $q \in S^1$ be a limit point of $f^{-1}(\gamma)$. Then the radial limit of f at q is p.*

Proof of Theorem 3.23. We may assume that $|f(z)| \leq 1$ and $f(z) \to 0$ along γ. Suppose $z_n \to \infty$ with

$$\pi - \alpha > \arg(z_n) > \alpha > 0.$$

We wish to show $|f(z_n)| \to 0$.

First suppose $|f(z)| \leq \epsilon$ on γ. If γ separates z_n from $\mathbb{R}_-$, then we have

$$\omega(z_n, \mathbb{H}, \gamma) \geq \omega(z_n, \mathbb{H}, \mathbb{R}_-) = \arg(z_n)/\pi > \alpha/\pi.$$

Hence $|f(z_n)| \leq \epsilon^p$, where $p = \alpha/\pi$. The same bound holds if γ separates z_n from $\mathbb{R}_+$, since $\pi - \arg(z_n) > \alpha$.

Now consider the general case. Given $\epsilon > 0$ there is an $R > 0$ such that $|f(z)| < \epsilon$ on γ when $|z| > R$. Then a change of coordinates of the form

$$z \mapsto z + R^2/z + r$$

sends $\mathbb{H} - B(0, R)$ to $\mathbb{H}$, and provides with a new path γ from 0 to ∞, and a new function $f(z)$ tending to zero on this path, now with $|f(z)| \leq \epsilon$ on γ. Since

$$\liminf \arg(z_n + R^2/z_n + r) = \liminf \arg(z_n) \geq \alpha,$$

and similarly for the limsup, we conclude by the preceding argument that $\limsup |f(z_n)| \leq \epsilon^p$. But ϵ was arbitrary, so $f(z_n) \to 0$. ∎

Example: Rectangles and bootstrapping. We conclude with two examples illustrating the polynomial and exponential decay of harmonic measure, respectively.

First we recall that

$$h(z) = \omega(z, \mathbb{H}, [-1, 1]) = \frac{\arg(z-1) - \arg(z+1)}{\pi}.$$

When $|z|$ is large, we have

$$h(z) \approx -\frac{2}{\pi}\frac{d}{dx}\arg(x + iy) = \frac{2}{\pi}\frac{y}{|z|^2}.$$

In particular, along the imaginary axis we have polynomial decay, $\omega(z) \asymp 1/y$.

Now consider, instead of $\mathbb{H}$, the half-infinite rectangle

$$S_+ = \{z = x + iy \ : \ x > 0 \quad \text{and} \quad y \in (0, \pi)\},$$

replace the interval $[-1, 1]$ with its base $i[0, \pi]$, let

$$u(z) = \omega(z, S_+, i[0, \pi]).$$

Here are two natural questions:

1. How rapidly does $u(z)$ tend to zero as $\operatorname{Re} z \to +\infty$?

2. For large x, what is the shape of the function $u(x + iy)$, $y \in [0, \pi]$?

To answer the first question, let

$$m(x) = \sup_{0 < y < \pi} u(x + iy).$$

We then find, by a simple nesting argument, that

$$m(x + y) \le m(x)m(y);$$

and by the maximum principle, $m(x) < 1$ for every $x > 0$. In particular, we have

$$m(nx) \le m(x)^n,$$

and thus the harmonic measure $u(x + iy) \to 0$ exponentially fast as $x \to \infty$.

40

For a more precise picture of the behavior of $u(z)$, we can use the conformal map $\cosh : S_+ \to \mathbb{H}$. Since $\cosh$ sends the base of the rectangle to $[-1, 1]$, we have

$$u(z) = h(\cosh z).$$

Now when x is large, we have

$$\cosh(x + iy) \approx \frac{e^x(\cos y + i \sin y)}{2},$$

and thus

$$u(x + iy) \approx \frac{4}{\pi} e^{-x} \sin y.$$

In particular, for any fixed $x \gg 0$, the function $u(x + iy)$ is approximately a multiple of $\sin(y)$.

4 Curvature

In this section we study the connection between the Laplacian and the Gaussian curvature of a conformal metric, and the idea of curvature as a measure. We then give an introduction to Ahlfors' theory of ultrahyperbolic metrics, and his 'elementary' proof of the Picard theorems.

The curvature of a metric. The Gaussian curvature of a metric on a surface $\Sigma \subset \mathbb{R}^3$ at a point p is given by $K = K_1 K_2$, where K_1 and K_2 are the two principal curvatures. Gauss's *theorema egregium* says that K is an intrinsic invariant of (Σ, g), i.e. it does not depend on the given embedding.

For a conformal metric $\rho = \rho(z)\,dz$, the curvature can be given by the formula

$$K(\rho) = \frac{-\Delta \log \rho}{\rho^2}.$$

The curvature is independent of choice of coordinates; that is, for any conformal mapping f,

$$K(f^* \rho) = f^* K(\rho),$$

since $\Delta \log |f| = 0$.

As an example, for the hyperbolic metric $\rho = |dz|/y$ on $\mathbb{H}$, we have

$$\Delta \log(1/y) = 1/y^2,$$

and hence $K(\rho) = -1$.

Curvature as a measure; Gauss–Bonnet. It is often natural to regard the curvature not as a function but as a measure, given by $K = K(\rho)\rho^2 = -\Delta \log \rho$. In Kähler geometry, the Ricci curvature of a metric is naturally a $(1, 1)$ form, which in the case of dimension one reduces to this measure.

The *Gauss–Bonnet theorem* relates the topology of a compact surface to its measure; it says:

$$2\pi\chi(\Sigma) = \int_{\Sigma} K + \int_{\partial\Sigma} k\, ds.$$

Here k is the geodesic curvature of the boundary. For example, if $D(r)$ is a Euclidean disk of radius r, then $K = 0$, $k = 1/r$, $\chi(D) = 1$, and we find

$$2\pi = \int_{S^1(r)} (1/r)ds.$$

For the unit sphere we have $\chi(S^2) = 1$, $K = 1$ and thus $\text{area}(S^2) = 4\pi$.

Another definition of curvature is that it records the turning of a vector under parallel transport. Using this idea, one can see that the curvature *measure* at a cone point of angle α should be a delta mass of size $\delta = 2\pi - \alpha$, the *angle defect*. Thus if we take Σ to be the double of a triangle with internal angles $(\alpha_1, \alpha_2, \alpha_3)$, we find

$$2\pi\chi(\Sigma) = 4\pi = \sum 2\pi - 2\alpha_i,$$

which gives $\sum \alpha_i = \pi$ as expected.

For a continuous version of this example, let Σ be the double of the unit disk. Then the curvature is concentrated on S^1 and proportional to arc length. In fact we have $K = 2\, d\theta$, so its total mass is 4π.

In fact, whenever Σ has boundary, if we double it across its boundary, we obtain curvature measure $2k\, ds$ along the crease. More generally, if we glue Σ_1 and Σ_2 together isometrically along their boundary, we obtain curvature $(k_1 + k_2)\, ds$ where they are joined. This is consistent with the fact that

$$\chi(\Sigma_1 \cup_{\partial} \Sigma_2) = \chi(\Sigma_1) + \chi(\Sigma_2),$$

since the boundary has Euler characteristic zero.

Ultrahyperbolic metrics. A smooth metric $\xi = \xi(z)\, |dz|$ is *ultrahyperbolic* if its curvature satisfies $K(\xi) \leq -1$.

The *hyperbolic* metric ρ_U on a plane region U is characterized by the fact that it is *complete* and that $K(\rho) = -1$, or equivalently:

$$\Delta \log \rho = \rho^2.$$

Similarly an ultrahyperbolic metric satisfies

$$\Delta \log \xi \geq \xi^2;$$

in addition, it does not have to be complete. For example, if $V \subset U$ then $\rho_U | V$ is an ultrahyperbolic metric.

(Note that any negatively curved metric has the property that $\log \rho$ is a *subharmonic* function; this is the complex analogue of a *convex* function.)

It is also convenient to allow an ultrahyperbolic to *vanish* at some points. At these points, no condition is impose on ξ.

Now is it obvious that if we make a metric smaller, we increase the size of its curvature. Thus for any $0 < r \leq 1$, the metric $r\rho_U$ is ultrahyperbolic on U.

Ahlfors made the important observation that the converse holds.

Theorem 4.1 *Let ξ be a conformal metric on Δ with $K(\xi) \leq K(\rho_\Delta) = -1$. Then $\xi \leq \rho_\Delta$.*

Proof. Let us first assume that ξ is bounded on the unit disk. Since $\rho_\Delta(z) = 2(1 - |z|^2)^{-1} \to \infty$ as $|z| \to 1$, the ratio ξ/ρ achieves its maximum at some point $p \in \Delta$. At the maximum, the second derivative must be non–negative in every direction, and hence

$$\Delta \log(\xi/\rho) \leq 0$$

at $z = p$. This gives

$$\xi^2 \leq \Delta \log \xi \leq \Delta \log \rho \leq \rho^2$$

at $z = p$, and hence $\xi(p) \leq \rho(p)$. Since the ratio is maximized here, we have $\xi \leq \rho$ throughout the disk.

For the general case, replace $\xi(z)$ with $r\xi(rz)$, $r < 1$, its pullback under $z \mapsto rz$, and pass to the limit as $r \to 1$. ∎

43

Corollary 4.2 *Let $f : \Delta \to U$ be an analytic map, and let ξ be an ultra-hyperbolic metric on U. Then f is weakly contracting from the hyperbolic metric on Δ to the ultrahyperbolic metric on U.*

Corollary 4.3 *If U admits an ultrahyperbolic metric, then any holomorphic map $f : \mathbb{C} \to U$ is constant.*

Supporting metrics. It is often useful to consider ultrahyperbolic metrics where $\xi(z)$ is just *continuous*, rather than smooth, just as one often considers *convex* functions that are not C^2.

In the convex case, the condition $f''(x) \geq 0$ is replaced by the condition that the distribution $f''(x)$ is a non-negative measure. Similarly, in the ultrahyperbolic case on requires that $\Delta \log \xi$ is a measure, bounded below by the measure $\xi^2 = \xi^2(z) |dz|^2$.

A convenient way of dealing with this technicality is to say that ξ is ultrahyperbolic at p if there is a locally defined *supporting metric* σ_p which is smooth, ultrahyperbolic, and satisfies $\sigma_p(p) = p$ and

$$\sigma_p(z) \leq \xi(z)$$

near p. With this definition, the proof of Ahlfors' Schwarz Lemma above goes through verbatim, using σ_p near p instead of ξ. The key point is that when ξ/ρ has a local maximum at p, so does σ_p/ρ, by the inequality above.

One can compare σ_p to the supporting hyperplane at p for the graph of a convex function.

An explicit metric. Ahlfors gave a very different proof of these two classical results using an ultrahyperbolic metric. His proof has the advantage of giving explicit lower bounds for ρ_U.

The first idea in the proof is that U is naturally divided into three regions, U_0, U_1 and U_∞. If we place the 3 punctures of U on the equator of the Riemann sphere, then U_p is just the set of points closest to p. Explicitly, we have

$$U_0 = \{z \ : \ \mathrm{Re}(z) \leq 1/2 \ \ \text{and} \ \ |z| \leq 1\}.$$

The desired ultrahyperbolic metric ξ will be given by an ordinary hyperbolic metric ρ_i on U_i. The construction will be symmetric enough that the pieces automatically fit together to give a continuous metric ξ.

The metric ρ_0 is defined as follows. Consider the Riemann mapping

$$f_0 : (\Delta, 0) \to (\widehat{\mathbb{C}} - [1, \infty], 0) = V_0.$$

This map can be normalized so that $f(1) = 1$; it is then given by

$$f(z) = \frac{4z}{(1+z)^2}.$$

(Note that f is just a rescaling of the Koebe function, whose image is $\mathbb{C} - [-\infty, -1/4]$.)

The map f_0 sends Δ^* bijectively to $V_0^* \supset U_0$. We define ρ_0 to be the pushforward of the hyperbolic metric, not on Δ^*, but on a large punctured disk; namely we use the metric

$$\rho_{\Delta^*(R)} = \frac{|dz|}{|z \log |R/z||} = \frac{|dz|}{|z|(\log(R) - \log|z|)}.$$

Note that the larger we take the value of R, the more this looks like a multiple of the cylindrical metric $|dz|/|z|$ when restricted to Δ^*.

Ahlfors shows that for a suitable choice of R, namely $R = e^4$, one can check that $d\rho_0/dx < 0$ along the line segment $\mathrm{Re}(z) = 1/2$ forming part of ∂U_0. It follows that ρ_0 provides a supporting metric for ξ on $\overline{U}_0$. A similar construction works on U_1 and U_∞, and hence ξ is an explicit ultrahyperbolic metric on U.

5 Capacity

In this section we study a measure of the size of a compact set $K \subset \mathbb{C}$ called its *capacity*. We will see this is the same as its *transfinite diameter*, and its *conformal radius* when K is connected.

The Vandermonde determinant. It is a basic identity in $\mathbb{Z}[a_1, \ldots, a_n]$ that

$$V(a_i) = \prod_{i<j}(a_j - a_i) = \det \begin{pmatrix} 1 & a_1 & a_1^2 & \ldots & a_1^{n-1} \\ 1 & a_2 & a_2^2 & \ldots & a_2^{n-1} \\ & & \ldots & & \\ 1 & a_n & a_n^2 & \ldots & a_n^{n-1} \end{pmatrix}.$$

The quantity $\Delta(a_i) = V(a_i)^2$ is invariant under permutations of the variables (a_i); it is the *discriminant* of the associated polynomial

$$P(z) = \prod_{1}^{n-1}(z - a_i).$$

45

One way to prove the identity above is to simply observe that the determinant vanishes whenever $a_i = a_j$, $i < j$.

Another is to interpret $V(a_i)$ as the determinant of the valuation map

$$E : \mathbb{C}[z]_{n-1} \to \mathbb{C}^n$$

that sends a polynomial $p(z)$ of degree $(n-1)$ to its values $(p(a_1), \ldots, p(a_n))$. The determinant is taken with respect to the monomial basis $(1, z, \ldots, z^{n-1})$. But any other sequence of *monic* polynomials $(p_0, \ldots, p_{n-1})$ is related to this basis by a triangular matrix with 1's along the diagonal. Hence

$$V(a_i) = \det[p_j(a_i)]. \tag{5.1}$$

Now take $p_0 = 1$, $p_1 = (z - a_1)$, $p_2(z) = (z - a_2)$, etc. Then $p_j(a_i) = 0$ when $j \geq i$, and we get a triangular matrix with diagonal entries

$$1, (a_2 - a_1), (a_3 - a_1)(a_3 - a_2), \ldots,$$

and the product of these entries is $V(a_i)$. It is useful to note that any other sequence of monic polynomials (p_i) will work as well.

The rational normal curve. The image of $\mathbb{P}^1$ in $\mathbb{P}^d$ under the map given in affine coordinates $r(t) = (t, t^2, \ldots, t^d)$ is called the *rational normal curve* of degree d. For example, R_2 is a conic; R_3 is a twisted cubic, etc.

Since R_d has degree d, its intersection with a hyperplane consists of at most d points. Thus any $(d+1)$ points on R_d are in general position (they span $\mathbb{P}^d$. This is equivalent to the statement that the Vandermonde determinant for distinct a_i is never zero. In fact, if we lift $r(a_i)$ to $v_i = (1, a_i, a_i^2, \ldots, a_i^d)$ in $\mathbb{C}^{d+1}$, then $V(a_i)$ is the (complex) volume of the parallelogram spanned by $(v_1, \ldots, v_{d+1})$ in $\mathbb{C}^{d+1}$.

Hadamard's inequality. The following inequality is well–known, and geometrically clear, for real matrices, but it is also true over the complex numbers: for any $n \times n$ matrix M, we have

$$|\det M| \leq \prod_1^n \|v_i\|,$$

where $v_i \in \mathbb{C}^n$ are the columns of M.

Transfinite diameter. We now turn to the study of a compact set $K \subset \mathbb{C}$. Let

$$D_n = \max_{z_1, \ldots, z_n \in K} \prod_{i < j} |z_j - z_i|.$$

There are $\binom{n}{2}$ terms in the product above; taking the geometric mean, we obtain the quantity $d_n(K) \geq 0$ characterized

$$d_n^{\binom{n}{2}} = D_n.$$

Note that $d_n(tK) = |t| d_n(K)$.

The *transfinite diameter* of K is defined by

$$d_\infty(K) = \lim_{n \to \infty} d_n(K).$$

To prove the limit exists we show:

Proposition 5.1 *We have $d_{n+1} \leq d_n$.*

Proof. Let D_{n+1}^i denote the product realizing D_{n+1} with the $(n+1)$ terms involving z_i eliminated. We then have $D_{n+1}^i \leq D_n$. Also, the term $|z_i - z_j|$ is eliminated from both D_{n+1}^i and D_{n+1}^j. Thus

$$D_{n+1}^{n-1} = \prod_1^{n+1} D_{n+1}^i \leq D_n^{n+1}.$$

The left hand side is d_{n+1} to the power $\binom{n+1}{2}(n-1)$, and the right hand side is d_n to the same power $\binom{n}{2}(n+1)$. ∎

Polynomial diameter. Another approach to measuring the diameter of K is to see how small one can make $P(K)$, where P ranges over *monic* polynomials $P_d(z)$ of degree d. Equivalently, we define

$$R_d(K) = \inf_{P_d} \max_K |P_d(z)|,$$

define $r_d(K) \geq 0$ so that
$$r_d^d(K) = R_d(K),$$

and let
$$r_\infty(K) = \lim_{d \to \infty} r_d(K) = \lim R_d(K)^{1/d}.$$

Noting that $t^d P_d(z/t)$ is monic when $P_d(z)$ is, we find that $r(tK) = |t| r(K)$.

Proposition 5.2 *The sequence $r_d(K)$ converges as $d \to \infty$.*

Proof. Let $P_d(z)$ be a monic polynomial of degree d that realizes $R_d(K)$. Fix $n > 0$ and consider a large index $N = nd + i$, $0 < i \leq n$. Since $P_n(z)^d P_i(z)$ is monic of degree N, we have

$$R_N \leq R_n^d R_i.$$

Taking the power $(1/N)$ on both sides and then the limit as $N \to \infty$, we have $R_i^{1/N} \to 1$ and $R_n^{d/N} \to R_n^{1/n} = r_n$. This shows that for each n,

$$\limsup r_N \leq r_n;$$

taking the $\liminf$ over n we find that the limit exists. ∎

Remark: Chebyshev polynomials. The quantity $r_\infty(K)$ is also called *Chebyshev's constant* (see e.g. [Gol, Ch. VII.1]). The Chebyshev polynomials, defined by

$$T_n(\cos\theta) = \cos(n\theta),$$

have degree n but are not monic; their leading coefficient is 2^{n-1}. They are extremal for $K = [-1, 1]$; thus $R_N(K) = 2^{1-n}$ and $r_\infty(K) = 1/2$.

One could also restrict to monic polynomials with all roots in K. The following proof shows this would not change the value of $r_\infty(K)$, and in fact we have just found another route to the transfinite diameter.

Theorem 5.3 *The polynomial diameter and transfinite diameter agree; that is, for all K, we have $r_\infty(K) = d_\infty(K)$.*

Proof. Let us first show that $r_\infty(K) \leq d_\infty(K)$. Suppose $V(a_1, \ldots, a_n) = D_n$. Consider the polynomial

$$Q_n(z) = (z - a_1) \cdots (z - a_n) \prod_{i<j} |a_j - a_i|.$$

Clearly Q_n/D_n is monic, and $\sup_K |Q_n(z)| \leq D_{n+1}$. Thus

$$R_n \leq \sup_K |Q_n(z)|/D_n \leq D_{n+1}/D_n.$$

Using the fact that $d_{n+1} \leq d_n$, we have

$$\frac{D_{n+1}}{D_n} \leq \frac{d_n^{\binom{n+1}{2}}}{d_n^{\binom{n}{2}}} = d_n^n.$$

48

Taking nth roots gives $r_n \leq d_n$ and hence $r_\infty(K) \leq d_\infty(K)$.

For the reverse inequality we use equation (5.1). Let $P_n(z)$ be a monic polynomial of degree n realizing R_n, and let $(a_1, \ldots, a_n)$ realize D_n. Then we can take $(1, P_1, \ldots, P_{n-1})$ as a basis for $\mathbb{C}[z]_{n-1}$, and we find

$$
D_n = \left| \det \begin{pmatrix}
1 & P_1(a_1) & P_2(a_1) & \ldots & P_{n-1}(a_1) \\
1 & P_1(a_2) & P_2(a_2) & \ldots & P_{n-1}(a_2) \\
 & & \ldots & & \\
1 & P_1(a_n) & P_2(a_n) & \ldots & P_{n-1}(a_n)
\end{pmatrix} \right|.
$$

Of course $|P_i(a_n)| \leq \max_K |P_i(z)| = R_i$. Thus the Euclidean norm of the ith column is at most $R_i \sqrt{n}$. The product of these norms is an upper bound for the determinant, and thus:

$$
D_n \leq n^{n/2} R_1 \cdots R_{n-1} = n^{n/2} r_1^1 r_2^2 \cdots r_{n-1}^{n-1}.
$$

Observe that $N = \binom{n}{2} = 1 + 2 + \cdots + (n-1)$ is the sum of the exponents of the terms r_i on the right. Taking the Nth root of both sides, we get

$$
d_n \leq n^{n/2N} (r_1 r_2^2 \cdots r_{n-1}^{n-1})^{1/N}.
$$

The power of n tends to one as $n \to \infty$, while the rest of the right hand side is a weighted geometric mean of $r_1, \ldots, r_{n-1}$. Since r_n converges, this shows:

$$
d_\infty(K) = \lim d_n \leq \lim r_n = r_\infty(K),
$$

as desired. $\blacksquare$

Example: the unit circle. The equality above is useful in estimating or proving $d_\infty(K)$ takes on a certain value, since concrete choices of points (a_i) provide a *lower bound* and concrete choices of polynomials $P_n(z)$ provide an *upper bound*. For example, one can show that

$$
d_\infty(S^1) = 1
$$

by bounding $d_n(S^1)$ from below using the nth roots of unity, and $r_n(S^1)$ from above using $P_n(z) = z^n$. (For the evaluation of d_n, one can use the fact that the columns of the Vandermonde determinant $|\det(\zeta_n^i)|$ are orthogonal to show $D_n = n^{n/2}$.)

Full sets. A compact set $K \subset \mathbb{C}$ is *full* if $\mathbb{C} - K$ is connected. Intuitively, K has no holes. For any compact set $J \subset \mathbb{C}$, one call fill in the holes to obtain a full set K, with

$$\partial K \subset \partial J \subset K.$$

If Ω is the unbounded component of $\mathbb{C} - J$, then $\Omega = \mathbb{C} - K$.

By the maximum principle, if K is full and P is a polynomial, then $\max_K |P| = \max_{\partial K} |P|$. This shows:

Proposition 5.4 *For any full compact set K, we have $r_\infty(K) = r_\infty(\partial K)$.*

It turns out that all the notions of diameter we consider in this section have the same property. It is thus useful to think primarily about the case where K is full.

The Green's function. We now turn to a third, equivalent method to measure the size of a compact set $K \subset \mathbb{C}$.

Let $\Omega \subset \mathbb{C}$ be a finitely–connected region. For each point $w \in \Omega$, there is a unique function $G(z, w)$ such that (i) $G(z, w) = 0$ for $z \in \partial\Omega$; and (ii) we have

$$\Delta_z G(z, w) = -2\pi\delta_w.$$

Equivalently, near w we have

$$G(z, w) = \log(1/|z - w|) + \phi(z),$$

where $\phi(z)$ is harmonic.

The Green's function is constructed by simply taking ϕ as above to be the solution to the Dirichlet problem with boundary values $\Phi(z) = \log|z - w|$.

The method of mirrors. The Green's function on the unit disk $\Omega = \Delta$ is easy to write down. First we observe that $G(z, 0) = \log(1/|z|)$. Then, we have use the fact that $G(z, w)$ is invariant under automorphisms: $G(Az, Aw) = G(z, w)$ for any $A \in \mathrm{Aut}(\Delta)$, together with the choice

$$A(z) = (z - w)/(1 - \overline{w}z),$$

to deduce that

$$
\begin{aligned}
G(z, w) &= G(Az, 0) = \log(1/|A(z)|) \\
&= \log|z - 1/\overline{w}| - \log|z - w| + \log|w|.
\end{aligned}
$$

Since $|A(z)| = 1$ on S^1, the values of $G(z, w)$ on the circle $|z| = 1$ are indeed zero. We note that

$$\Delta_z G(z, w) = 2\pi(\delta_{1/\overline{w}} - \delta_w);$$

we have placed a mirror charge outside the disk at $1/\overline{w}$, which is the reflection of w through the circle. The symmetry of these two charges causes the potential to be constant along S^1.

Convolution with G. The Green's function allows one to the solve inhomogeneous equation

$$\Delta\phi = -2\pi\psi$$

on Ω, subject to the boundary condition $\phi|\partial\Omega = 0$. The solution, which is unique, is given simply by

$$\phi(z) = G * \psi = \int_\Omega G(z, w)\psi(w)\,|dw|^2.$$

The boundary condition, that ϕ is constant on $\partial\Omega$, arises naturally in electrostatics when the boundary is a conductor. Because electrons can move freely in a metal, they flow along the boundary until the potential is the same everywhere.

Symmetry. For $\Omega = \Delta$, we have seen that $G(z, w) = G(w, z)$; this turns out to be a general property of the Green's function. Equivalently, G is *self–adjoint* as a convolution operator. This symmetry arises because formally $G = \Delta^{-1}$, and the Laplacian is self–adjoint.

In terms of an orthonormal basis of eigenfunctions for the Euclidean Laplacian, satisfying

$$\Delta f_n = \lambda_n f_n,$$

and subject to the boundary conditions $f_n|\partial\Omega = 0$, the Green's function can be expressed as

$$G(z, w) = \sum_n \lambda_n^{-1} f_n(z) f_n(w).$$

This formula also shows $G(z, w) = G(w, z)$.

Bessel functions. We remark that the radial eigenfunctions of the Euclidean Laplacian on the unit disk can be expressed in terms of solutions to Bessel's differential equation. The first point is that the Laplacian in polar coordinates (r, θ) is given by

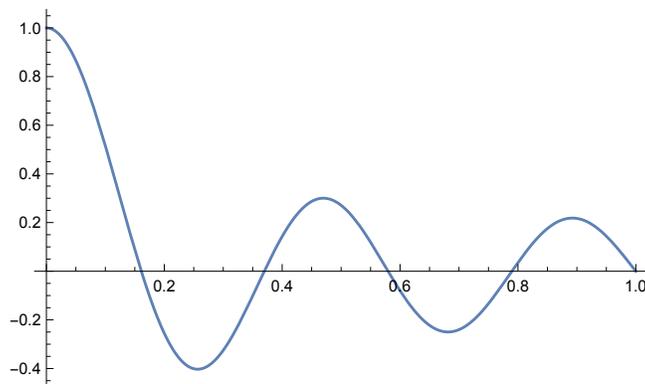$$\Delta\phi = \frac{d^2\phi}{dr^2} + \frac{1}{r}\frac{d\phi}{dr} + \frac{1}{r^2}\frac{d^2\phi}{d\theta^2}.$$

Figure 6. The radial Bessel eigenfunction $J_0(r\alpha_5)$.

To check this, observe that

$$
\begin{aligned}
(\Delta\phi)r\,dr\,d\theta &= d*d\phi = d*(\phi_r\,dr + \phi_\theta\,d\theta) = d(\phi_r r\,d\theta - \phi_\theta r^{-1}\,dr) \\
&= (\phi_{rr}r + \phi_r + \phi_{\theta\theta}r^{-1})\,dr\,d\theta.
\end{aligned}
$$

Thus the eigenvalue equation $\Delta\phi = -\lambda\phi$ for a radially symmetric function

$$
\frac{d^2\phi}{dr^2} + \frac{1}{r}\frac{d\phi}{dr} + \lambda\phi = 0.
$$

For $\lambda = 1$ this is Bessel's equation; it has a unique solution $J_0(r)$ with the property that $J_0(0) = 1$. (The other standard solution, $Y_0(r)$, blows up at $r = 0$.) The zeros of $J_0(r)$ are at $\alpha_1 < \alpha_2 < \cdots$. To obtain functions that vanish on the circle, one considers

$$
\phi_n(r) = J_0(\alpha_n x);
$$

these satisfy

$$
\Delta\phi_n = -\alpha_n^2\phi_n,
$$

and give an eigenbasis for the Laplacian on radially symmetric functions.

Note the similarity to the eigenfunctions $\phi_n(x) = \cos(\beta_n x)$ of the Laplacian on $[0,1]$ with Dirichlet boundary conditions, with $\beta_n = 2\pi(n + 1/2)$; these are also obtained from the single function $\cos(x)$ by rescaling the x–coordinate. In fact Bessel's equation for large r is well–approximated by $d^2\phi/dr^2 + \phi = 0$, which has trigonometric solutions.

52

A complete eigensystem for Δ on the unit disk can be obtained using separation of variables, i.e. by considering functions such as $\phi_n(r)\sin(m\theta)$.

The exterior Green's function. Now suppose $\Omega = \mathbb{C} - K$ is the complement of a compact set $K \subset \mathbb{C}$ made up of finitely many disjoint Jordan disks. We can then associated to Ω the Green's function $G(z)$ for $w = \infty$. It is the unique continuous function on $\overline{\Omega}$ satisfying $\Delta G = 0$, $G|\partial\Omega = 0$, and

$$G(z) = \log|z| + \gamma + O(1/|z|) \tag{5.2}$$

near infinity. It is natural to extend G to 0 on all of K.

The equilibrium measure. Associated to the Green's function at infinity is the *equilibrium measure* μ on K, characterized by

$$2\pi\mu = \Delta G.$$

When $\partial\Omega$ is smooth, we have

$$2\pi\mu = *dG|\partial\Omega;$$

equivalently, $\mu = (1/2\pi)|\nabla G|ds$.

This charge distribution μ has the remarkable property that its associated potential, $\phi = E * \mu$, is constant on K. Indeed, by the behavior at infinity (equation (3.2), we have

$$\phi(z) = (E * \mu)(z) = G(z) - \gamma,$$

and thus $\phi|K = -\gamma$ on K.

Capacity. We define the *capacity* of the compact set K in terms of equation (5.2) by

$$\mathrm{cap}(K) = \exp(-\gamma).$$

For example, if $K = \overline{\Delta}(R)$, then $G(z) = \log|z/R|$, $\gamma = -\log(R)$ and $\mathrm{cap}(K) = R$.

A general compact set K, with $\Omega = \mathbb{C} - K$ connected, can be written as $K = \bigcap K_n$, where $K_1 \supset K_2 \supset \cdots$ and each K_i is a finite union of disjoint, smoothly bounded disks. We then set

$$\mathrm{cap}(K) = \lim \mathrm{cap}(K_n).$$

Capacity and the Riemann mapping theorem. The case where K is a connected, full set is especially geometric. In this case we have:

Theorem 5.5 *Suppose we have a conformal isomorphism*

$$f : (\mathbb{C} - K) \to (\mathbb{C} - \overline{\Delta}(R)),$$

with $f(z) = z + O(1)$ near infinity. Then

$$\mathrm{cap}(K) = R.$$

In fact, in this case the Green's function for K is given simply by $G(z) = \log |f(z)/R|$, so $\gamma = -\log R$.

The Riemann mapping also gives a good geometric picture of the equilibrium measure. Recall that when K is locally connected, the Riemann map extends to a continuous function

$$f : S^1 \to \partial K.$$

Corollary 5.6 *Provided K is locally connected, its equilibrium measure satisfies*

$$\mu(A) = |f^{-1}(A)|/2\pi$$

for any Borel set $A \subset K$.

Here the right–hand side is the normalized arclength of $f^{-1}(A) \subset S^1$.

Examples. We have $\mathrm{cap}(\overline{\Delta}) = 1$ and $\mathrm{cap}(\overline{\Delta}(R)) = R$. The map $f(z) = z + 1/z$ shows that

$$\mathrm{cap}([-2, 2]) = 1,$$

and more generally

$$\mathrm{cap}([a, b]) = (1/4)|a - b|.$$

Capacity and transfinite diameter. We will now see that all three of our methods for measuring the size of a compact set $K \subset \mathbb{C}$ agree:

Theorem 5.7 *For any compact set $K \subset \mathbb{C}$, its capacity K and its transfinite diameter agree: we have $\mathrm{cap}(K) = d_\infty(K)$.*

For convenience we will assume K is a union of disjoint disks with smooth boundaries, and normalize (by scaling) so that

$$\mathrm{cap}(K) = 1.$$

Let $E(z) = \log |z|$ as earlier. Then the potential of its equilibrium measure μ recovers the Green's function of K: we have

$$G = E * \mu.$$

The general case follows by a limiting argument.

Lemma 5.8 *Under these assumptions, the potential $\psi = E * \nu$ of any finite measure ν on K satisfies*

$$\inf_{\partial K} \psi(z) \leq 0 \leq \sup_{\partial K} \psi(z).$$

Proof. We may assume ν is a probability measure. The difference $\psi - G$ is then harmonic at infinity, and vanishes there. By the maximum principle, 0 lies between the maximum and minimum of $\psi - G$ on ∂K. But $G = 0$ on ∂K. ∎

Proof of Theorem 5.7. We will show $\mathrm{cap}(K) = r_\infty(K)$, the equality $r_\infty(K) = d_\infty(K)$ having already been established as Theorem 5.7.

Let $P_n(z)$ be a monic polynomial of degree n with zeros $a_1, \ldots, a_n$. Let $\nu = 1/n \sum \delta_{a_i}$. Then

$$\psi(z) = (1/n) \log |P_n(z)| = E * \nu$$

satisfies $\sup_K \psi(z) \geq 0$ by the preceding lemma. Hence $\sup_K |P_n(z)| \geq 1$ and we have $r_\infty(K) \geq 1$ as well.

For the reverse inequality, let μ_n be a sequence of atomic probability measures converging weakly to μ, supported on K, each with equal masses at n distinct points. We then have a monic polynomial $P_n(z)$ such that the potential of μ_n is given by

$$\phi_n = (1/n) \log |P_n(z)| = E * \mu_n.$$

We also have $\phi_n \to \phi$ uniformly on compact subsets of $\Omega = \mathbb{C} - K$. Given $\epsilon > 0$, let $D = \phi^{-1}[0, \epsilon]$. Note that D is compact, $\phi|\partial D = \epsilon$, and K is contained in the interior of D. Thus $\phi_n \to \phi$ uniformly on ∂D, and hence $|\phi_n| < 2\epsilon$ on ∂D for all $n \gg 0$. Applying the maximum principle to $\phi_n|D$, and noting that $\phi_n(a_i) = -\infty$, we conclude that $\phi_n(z) \leq 2\epsilon$ on D for all $n \gg 0$.

It follows that $\sup_K |P_n(z)|^{1/n} \leq \exp(2\epsilon)$ for all $n \gg 0$, and hence $r_\infty(K) \leq 1$.

The combination of these inequalities gives $r_\infty(K) = 1 = \mathrm{cap}(K)$. ∎

Koebe 1/4, *reprise*. The equality $d_\infty(K) = \mathrm{cap}(K)$ reveals some monotonicity properties of capacity that are not obvious from its definition using potential theory. For example, we have:

Theorem 5.9 *Let $K \subset \mathbb{C}$ be compact and let $K' \subset \mathbb{R}$ be its vertical projection to the real axis. Then $\mathrm{cap}(K') \leq \mathrm{cap}(K)$.*

Proof. Projection decreases distances, so $d_\infty(K') \leq d_\infty(K)$. ∎

Corollary 5.10 *If K is connected, then $\mathrm{cap}(K) \geq \mathrm{diam}(K)/4$.*

Proof. A suitable projection of K is an line segment of length $\mathrm{diam}(K)$, and the capacity of a segment I is $|I|/4$. ∎

Corollary 5.11 *If $f \in S$, and $p \notin f(\Delta)$, then $|p| \geq 1/4$.*

Proof. Let $L = \widehat{\mathbb{C}} - f(\Delta)$. Then $K = 1/L$ has capacity one by the normalization of f, and it contains $z = 0$ and $1/p$. Thus $|1/p| \leq \mathrm{diam}(K) \leq 4$ by the preceding Corollary, and hence $|p| \geq 1/4$. ∎

Spanning more than 2 points. The discussion above implies that $\mathrm{cap}(K)$ is minimized, among all connected compact sets K containing $P = a, b$, by the line segment $[a, b]$.

What happens for other finite sets P? The *(Polya–)Chebotarev problem* is to minimize $\mathrm{cap}(K)$ among all continua containing K. It has a unique solution. The optimal set K is a topological tree; it forms the singular trajectories of a suitable meromorphic quadratic differential with a double pole at infinity and simple poles at the points in P [Str]. For an explicit solution in the case $|P| = 3$, see [Sch].

Remarks: Energy, harmonic measure and the Riemann maps.

1. Capacity can also be discussed in terms of the *energy* of a probability measure on K, given by:

$$I(\nu) = \int_{K \times K} \frac{1}{\log|z - w|}\, \nu(z)\, \nu(w).$$

Physically, $I(\nu)$ measures the energy required to bring a unit charge distribution 'from infinity' into the configuration ν; energy is required because the charges repel one another. (In fact this energy is infinite; thus $I(\nu)$ is more of a 'renormalized energy', that is allowed to assume negative values.) We then have:

Proposition 5.12 *The equilibrium measure μ on K uniquely minimizes the energy $I(\nu)$ over all probability measures supported on K; and*

$$\mathrm{cap}(K) = \exp(-I(\mu)).$$

(Note that it takes more energy to confine a unit charge to a smaller region.) Let us check the final formula: assuming $\mathrm{cap}(K) = 1$, we have $G(z) = \mu * \log|z|$, and $G(z) = 0$ on K, so

$$I(\mu) = -\int_K G(w)\,\mu(w) = 0.$$

For more details, see [Ah2, Ch. 2-2].

2. In dimension 3, the harmonic potential $\log(1/r)$ of a point charge is replaced by $1/r$. With that change, $I(\mu)$ is always positive; it takes only a finite amount of energy to bring a unit charge from infinity to a bounded region. A related difference between dimensions two and three, as we will later see, is that a random walk on $\mathbb{Z}^2$ is recurrent, while a random walk on $\mathbb{Z}^3$ tends to infinity.

3. For $\Omega = \widehat{\mathbb{C}} - K$ we have:

Proposition 5.13 *The equilibrium measure μ on K coincides with harmonic measure on $\partial\Omega$ as seen from $z = \infty$. That is, for any Borel set $A \subset K$, we have $\mu(A) = \omega(\infty, \Omega, A)$.*

This is clear when K is a Jordan disk, using the Riemann mapping theorem. The general case can be established using $I(\mu)$.

4. The proof that $r_\infty(K) = \mathrm{cap}(K)$ can be adapted to show — when K is a Jordan disk and $\mathrm{cap}(K) = 1$ — that there exists monic polynomials $P_n(z)$, with all zeros in K, such that

$$\left(\max_K |P_n|\right)^{1/n} \to r_\infty(K) \quad \text{and} \quad P_n(z)^{1/n} \to f(z),$$

where $f : (\mathbb{C} - K) \to (\mathbb{C} - \overline{\Delta})$ is the Riemann mapping, normalized so that $f(z) = z + O(1)$ near infinity. Indeed, the proof shows that P_n can be chosen so that $\log|P_n(z)| \to \log|f(z)|$ uniformly on compact subsets of $\mathbb{C} - K$. Special cases of this result arise from dynamics, as we will see below.

Julia sets and capacity. Very complicated sets whose capacity can be computed exactly arise naturally in complex dynamics.

Let $f(z) = z^d + a_1 z^{d-1} + \cdots + a_d$ be a monic polynomial of degree $d > 1$. The *filled Julia set* of f is defined by

$$K(f) = \{z \,:\, \sup|f^n(z)| < \infty\},$$

where $f^n(z)$ denotes the nth iterate of $f(z)$. It is not hard to show that $K(f)$ is compact and its complement is connected. Another nice fact is that $K(f)$ itself is connected if and only if it contains all the critical points of f.

When $K(f)$ is connected, there is a simple formula for the Riemann mapping

$$F : (\mathbb{C} - K(f)) \to (\mathbb{C} - \overline{\Delta}),$$

namely

$$F(z) = \lim_{n\to\infty} (f^n(z))^{1/d^n}.$$

Of course some care must be taken in extracting the d^nth root; it is chosen so that $F(z) = z + O(1)$ for $|z| \gg 1$. This map satisfies the functional equation

$$F(f(z)) = F(z)^d,$$

i.e. outside of $K(f)$, it conjugates $z \mapsto f(z)$ with the simpler dynamical system $z \mapsto z^d$.

Since the Green's function for $K(f)$ is given by $\log|F(z)|$, it too has a dynamical expression:

$$G(z) = \lim d^{-n} \log^+ |f^n(z)|.$$

Here $\log^+(x) = \max(0, \log x)$. Note that $G(z) = 0$ on $K(f)$. It is easy to see that $G(z) = \log|z| + O(1/|z|)$ for $|z| \gg 1$. Moreover this expression makes sense even when $K(f)$ is disconnected. Summing up, we have:

Theorem 5.14 *Let $f(z)$ be a monic polynomial of degree $d > 1$. Then its filled Julia set $K(f)$ has capacity one.*

6 Extremal length

In this section we introduce the method of extremal length, and use it to study the study the moduli of quadrilaterals and annuli. We start by classifying these two types of regions, whose moduli are given by a single real number.

The classification of annuli. An *annulus* $A \subset \widehat{\mathbb{C}}$ is a region such that $\widehat{\mathbb{C}} - A$ has two connected components. This is equivalent to the condition that $\pi_1(A) \cong \mathbb{Z}$.

Aside from the punctured disk and plane, Δ^* and $\mathbb{C}^*$, one has the standard family of annuli

$$A(R) = \{z \ : \ 1 < |z| < R\}.$$

Every annulus is isomorphic to one of these. In the metric $|dz|/|z|$, $A(R)$ becomes a right Euclidean annulus of height $h = \log R$ and circumference $c = 2\pi$. Its *modulus* is defined by

$$\operatorname{mod} A(R) = \frac{h}{c} = \frac{\log R}{2\pi}.$$

Theorem 6.1 *Any annulus $A \subset \mathbb{C}$ is conformally isomorphic to $\mathbb{C}^*$, Δ^* or $A(R)$ for a unique $R > 1$.*

Proof. Normalize by a Möbius transformation so that A separates 0 from ∞, and let $S = \log(A)$ be its preimage under the covering map $\exp : \mathbb{C} \to \mathbb{C}^*$. Then S is simply–connected, and $A \cong S/\langle h \rangle$, where $h(z) = z + 2\pi i$.

If $S = \mathbb{C}$ then $A \cong \mathbb{C}^*$. Otherwise, $S \cong \mathbb{H}$ by the Riemann mapping theorem, and h can be transported to a fixed–point free Möbius transformation $g \in \operatorname{Aut}(\mathbb{H})$, such that

$$A \cong \mathbb{H}/\langle g \rangle.$$

Normalize so that $g(\infty) = \infty$. If g is parabolic, then we can further normalize so that $g(z) = z + 1$; then $q(z) = \exp(-2\pi i z)$ gives an isomorphism between $\mathbb{H}/\langle g \rangle$ and Δ^*.

Otherwise, g is hyperbolic, and we can normalize so $g(z) = \lambda z$, $\lambda > 1$. Then for a suitable choice of α,

$$f(z) = z^{i\alpha} = \exp(i\alpha \log z)$$

gives an isomorphism between $\mathbb{H}/\langle g \rangle$ and a round annulus, $1/R < |z| < 1$. One should take $\alpha = 2\pi/\log \lambda$, so that $f(\lambda z) = f(z)$; and then

$$1/R = f(-1) = \exp(-\pi\alpha) = \exp(-2\pi^2/\log \lambda).$$

Since the hyperbolic on $A(R)$ uniquely determines the length $L = \log \lambda$ of its core geodesic, $A(R) \cong A(S)$ iff $R = S$. ∎

Here is another proof of the last step: if $f : A(R) \cong A(S)$ is a conformal map, normalized so $f(S^1) = S^1$, then by repeated Schwarz reflection it can be extended to a map $f : \mathbb{C}^* \to \mathbb{C}*$. Thus $f(z) = az$ for some $a \neq 0$, and $|a| = 1$ by our initial normalization, so $R = S$.

The modulus of an annulus. If $A \cong A(R)$ we define $\mathrm{mod}(A) = \mathrm{mod}(A(R))$. The proof above shows that

$$\mathrm{mod}(A) = \frac{\log R}{2\pi} = \frac{\pi}{\mathrm{length}(\gamma)},$$

where γ is the unique closed hyperbolic geodesic in A. (For $A = A(R)$, γ is the circle of radius $\sqrt{R}$.)

We adopt the convention that $\mathrm{mod}(\Delta^*) = \mathrm{mod}(\mathbb{C}^*) = \infty$.

The classification of quadrilaterals. A *quadrilateral* is a Jordan domain $Q \subset \mathbb{C}$ with 4 distinct distinguished points (q_1, q_2, q_3, q_4) in its boundary, whose cyclic ordering is compatible with the standard orientation of ∂Q.

An *isomorphism* between quadrilaterals is a conformal map $f : Q \to Q'$ preserving the distinguished points. It is not hard to see that there is an automorphism of Q sending (q_1, q_2, q_3, q_4) to (q_3, q_4, q_1, q_2).

There are 2 standard models for a quadrilateral. The first is $Q(\lambda) = \mathbb{H}$ with the point $(0, 1, \lambda, \infty)$ distinguished, $\lambda > 1$. The second is the rectangle

$$R(a, b) = \{z = x + iy \; : \; (x, y) \in (0, a) \times (0, b)\},$$

with the sides points $(0, a, a + ib, ib)$ distinguished.

The *modulus* of $R(a, b)$ is given by

$$\mathrm{mod}\, R(a, b) = \frac{a}{b}.$$

Note that if we glue the horizontal sides of $R = R(h, 2\pi)$ together, we obtain an annulus with $\mathrm{mod}(A) = \mathrm{mod}(R)$. We also note that, by a Schwarz reflection argument, we have

$$R(a, b) \cong R(a', b') \iff a/b = a'/b'.$$

60

Theorem 6.2 *Every quadrilateral is isomorphic to a unique quadrilateral of the form $Q(\lambda)$, and to a unique rectangle of the form $R(a, b)$ with $ab = 1$.*

Proof. The first statement is immediate from the Riemann mapping theorem and the fact that $\text{Aut}(\mathbb{H})$ can be used to move any 3 points in $\partial\mathbb{H}$ to $(0, 1, \infty)$.

For the second statement, consider the holomorphic 1–form

$$\omega = \frac{dz}{\sqrt{z(z-1)(z-\lambda)}} \tag{6.1}$$

in $\mathbb{H}$. Any solution to $df = \omega$ maps $\mathbb{H}$ to an open rectangle R, sending $(0, 1, \lambda, \infty)$ to its vertices. ∎

Calculations with elliptic functions. It is a remarkable fact, implicit in the proof above, that the modulus of $Q(\lambda)$ can be computed exactly. Namely, if we let

$$a(\lambda) = \int_0^1 \omega \quad \text{and} \quad b(\lambda) = \int_1^\lambda \omega, \tag{6.2}$$

where ω is given by equation (6.1), then $Q(\lambda)$ has modulus $|a(\lambda)/b(\lambda)|$. In fact we have an isometry between the spaces $(\mathbb{H}, |\omega|)$ and $(R(a, b), |dz|)$.

Extremal length. Let $\Omega \subset \mathbb{C}$ be a region. A *curve* is a continuous path $\gamma : [a, b] \to \Omega$. A *conformal metric* $\rho = \rho(z)\,|dz|$ on Ω is specified by a Borel measurable function $\rho(z) \geq 0$. Provided γ is rectifiable, e.g. Lipschitz, its length in the metric ρ is defined by

$$L(\gamma, \rho) = \int_\gamma \rho = \int_a^b \rho(\gamma(t))\,|\gamma'(t)|\,dt.$$

We think of the elements of the path family as competitors for *geodesics* or shortest paths, and set

$$L(\Gamma, \rho) = \inf_{\gamma \in \Gamma} L(\gamma, \rho).$$

The *area* of Ω in the same metric is given by:

$$A(\Omega, \rho) = \int_\Omega \rho^2 = \int_\Omega \rho(z)^2\,|dz|^2.$$

Both the length and area may be infinite.

The *extremal length* of a collection of curves Γ in Ω is defined by

$$\lambda(\Gamma, \Omega) = \sup_\rho \frac{L(\Gamma, \rho)^2}{A(\Omega, \rho)}.$$

The supremum is taken over all metrics that give Ω positive area.

Extremal length for quadrilaterals. Extremal length provides a natural approach to the moduli of quadrilaterals and annuli, and to other conformal invariants.

To illustrate this connection, let (Q, q_1, q_2, q_3, q_4) be a quadrilateral. The a–sides of Q are the arcs $[q_1, q_2]$ and $[q_3, q_4]$ in ∂Q; the b–sides are defined similarly. Let Γ_a denote the paths in Q connecting the b–sides, or equivalently separating the a–sides. Let $\Gamma_b(Q)$ denote those connecting the a–sides.

Theorem 6.3 *We have $\lambda(\Gamma_a, R(a, b)) = a/b = \mathrm{mod}(R(a, b))$.*

Proof. For the Euclidean metric on $R = R(a, b)$, we have $L(\Gamma_a, |dz|) = a$ and $A(R, |dz|) = ab$, so $\lambda(\Gamma_a) \geq a/b$.

To prove the reverse inequality, let ρ be any conformal metric on R. Let γ_y be the horizontal line segment at height y in R. We then have

$$\int_R \rho(z) \, |dz|^2 = \int_0^b L(\gamma_y, \rho) \, dy \geq bL(\Gamma_a, \rho).$$

By the Cauchy–Schwarz inequality, we also have

$$\left(\int_R \rho \right)^2 \geq \int \rho^2 \int 1^2 = A(R, \rho)ab.$$

This shows that
$$A(R, \rho)ab \geq b^2 L(\Gamma_a, \rho)^2,$$

which gives
$$\lambda(\Gamma_a) = \sup_\rho \frac{L(\Gamma_a, \rho)^2}{A(R, \rho)} \leq \frac{a}{b},$$

completing the proof. ∎

62

Since extremal length is a conformal invariant, we have:

Corollary 6.4 *For any quadrilateral Q, we have*

$$\mathrm{mod}(Q) = \lambda(\Gamma_a, Q)) = 1/\lambda(\Gamma_b, Q)).$$

Extremal length and annuli. Let A be an annulus. Let Γ_r denote the paths in A joining its two boundary components, and Γ_θ the loops in A separating them. Similar reasoning, using the metric $|dz|/|z|$ on $A(R)$, gives:

Theorem 6.5 *For any annulus A, we have*

$$\mathrm{mod}(A) = \lambda(\Gamma_r, A) = 1/\lambda(\Gamma_\theta, A).$$

Besikovitch's theorem. Here is a result in elementary geometry that can be efficiently proved used extremal length.

Theorem 6.6 *Let $Q \subset \mathbb{C}$ be a Jordan domain with ∂Q comprised of cyclically ordered arcs $\alpha, \beta, \alpha', \beta'$. Then:*

$$\mathrm{area}(Q) \geq d(\alpha, \alpha') \cdot d(\beta, \beta').$$

Proof. Make Q into a quadrilateral so its a–sides are α, α' and its b–sides are β, β'. Let $\rho = |dz|$ be the Euclidean metric. Then $A(Q, \rho) = \mathrm{area}(Q)$,

$$d(\alpha', \alpha) = L(\Gamma_b, \rho), \quad \text{and} \quad d(\beta, \beta') = L(\Gamma_a, \rho).$$

Since

$$1 = \lambda(\Gamma_a)\lambda(\Gamma_b) \geq \frac{L(\Gamma_a, \rho)^2 L(\Gamma_b, \rho)^2}{A(Q, \rho)^2},$$

the theorem follows. ∎

Bounds. Any metric ρ on Q gives a lower bound for the extremal length of both Γ_a and Γ_b. Thus any metric gives upper and lower bounds for $\mathrm{mod}(Q)$. The same is true for the path families Γ_r and Γ_θ.

For example, the Euclidean metric on $A = A(R)$ give $\mathrm{area}(A) = \pi(R^2 - 1)$, $L(\Gamma_r) = R - 1$ and $L(\Gamma_\theta) = 2\pi$. Thus

$$\frac{(R-1)^2}{\pi(R^2-1)} = \frac{R-1}{\pi(R+1)} \leq \mathrm{mod}(A(R)) \leq \frac{\pi(R^2-1)}{4\pi^2} = \frac{R^2-1}{4\pi}.$$

63

These bounds are close to the actual value, $\text{mod}(A(R)) = \log R/(2\pi)$, when R is close to 1, since then the Euclidean metric is a good approximation to the optimal metric $|dz|/|z|$. For large R they are way off: the lower bound tends to a constant, while the upper bound grows like $R^2 \gg \log R$.

Beurling's criterion for an extremal metric. (See [Ah2, §4.7].) A metric is *extremal* if it realizes the supremum in the definition of $\lambda(\Gamma)$.

Theorem 6.7 *Suppose the area measure $\rho^2|\Omega$ is swept out by the linear measures on ρ–geodesics:*

$$\{\rho|\gamma \ : \ \gamma \in \Gamma \ and \ L(\gamma, \rho) = L(\Gamma, \rho)\}.$$

Then ρ is extremal for Γ.

There are two ways to make the hypothesis 'swept out' precise. One is to require that there is a measurable family of geodesics γ_t, and a weight function $w(t) \geq 0$, such that

$$\int_A \rho^2 = \int_0^1 w(t) L(\gamma_t \cap A) \, dt$$

for all Borel sets $A \subset \Omega$. (This is an 'integral geometry' condition.)

Another is to require that, for all Borel functions h, whenever $\int_\gamma h\rho \geq 0$ for all geodesics γ, we have $\int h\rho^2 \geq 0$ as well.

We will use the first to explain the proof.

Proof. Consider any other Borel metric α. We may assume both α and ρ are normalized to give Ω area one.

Now for any ρ–geodesic $\gamma \in \Gamma$, we have

$$L(\Gamma, \alpha) \leq L(\gamma, \alpha) = \int_\gamma (\alpha/\rho)\rho = \langle \alpha/\rho, \rho|\gamma \rangle$$

where the last expression is the pairing between functions and measures. Since the probability measure ρ^2 is a limit of convex combinations of the probability measures $(\rho|\gamma)/L(\Gamma, \rho)$, we have

$$\frac{L(\Gamma, \alpha)}{L(\Gamma, \rho)} \leq \langle \alpha/\rho, \rho^2 \rangle = \int_\Omega \alpha\rho \leq \left(\int \alpha^2 \int \rho^2 \right)^{1/2} = 1$$

by Cauchy-Schwarz. Thus $L(\Gamma, \alpha) \leq L_(\Gamma, \rho)$, and therefore ρ maximizes the ratio of length–squared to area. ∎

Examples.

1. *Annuli and quadrilaterals.* The area form associated to $|dz|$ on a rectangle $R(a, b)$ can be expressed as $dx\, dy$, which is a continuous linear combination of the length measures dx along horizontal geodesics. Thus Beurling's theorem implies this metric is extremal, as we have already seen directly. The same is true for the cylindrical metric $|dz|/|z|$ on an annulus $A(R)$.

2. *Simple curves on a torus.* Consider the torus $X = \mathbb{C}/\mathbb{Z} \oplus \mathbb{Z}\tau$, $\operatorname{Im}\tau > 0$. Let Γ be the family of loops on X in the homotopy class of $[0, 1]$. By Beurling's criterion, the flat metric $|dz|$ on X is extremal. The area of X is $\operatorname{Im}\tau$ and the length of the geodesic $[0, 1]$ is 1, so we find

$$\lambda(\Gamma) = \frac{1}{\operatorname{Im}\tau}.$$

3. *A triangle.* Let $T \subset \mathbb{C}$ be an equilateral triangle with a base of length $b = 2$ and height $h = \sqrt{3}$. Let Γ be the set of paths that touch all three sides of T. For the Euclidean metric, a reflection argument shows that 'billiard paths' are geodesics of length h. Also area measure is an average of billiard paths. Thus ρ is extremal, and we find

$$\lambda(\Gamma) = \frac{h^2}{(1/2)bh} = h = \sqrt{3}.$$

   Note that all triangles (Jordan domains with 3 marked boundary points) are conformally equivalent,

4. *The real projective plane.* Let $X = \mathbb{RP}^2$. Then X is canonically a Riemann surface, in the sense that its universal cover S^2 has a unique conformal structure, and this structure is preserved (up to orientation) by the deck transformations of S^2/X.

   Let Γ be the family of all loops generating $\pi_1(X) \cong \mathbb{Z}/2$. We claim $\lambda(\Gamma) = \pi/2$.

   To see this, let ρ be the round metric on X, making its universal cover S^2 into the sphere of radius 1. If we average linear measure on a great circle over the rotation group of S^2, we obtain an invariant measure which must be the usual area form. Thus ρ satisfies Beurling's criterion. The minimal length of a curve joining antipodal points on S^2 is π, and the area of $\mathbb{RP}^2$ is 2π, so $\lambda(\Gamma) = \pi^2/(2\pi) = \pi/2$.

Addition theorems. An inclusion between annuli, $A \subset A'$, is *essential* if A separates the boundary components of A'; equivalently, if $\pi_1(A)$ maps injectively into $\pi_1(A') \cong \mathbb{Z}$.

Theorem 6.8 *Let $B_1, B_2 \subset B$ be a pair of disjoint essential annuli. Then:*

$$\mathrm{mod}(B) \geq \mathrm{mod}(B_1) + \mathrm{mod}(B_2).$$

Proof. Let ρ_i be extremal metrics on B_i, giving length and area L_i and B_i for the path family $\Gamma_r(B_i)$. Scale these metrics so $A_i = L_i$, and let $\rho = \rho_1 + \rho_2$. Let $\Gamma = \Gamma_r(B)$ have length and area L and B in the metric ρ. Any curve $\gamma \in \Gamma$ must cross both B_1 and B_2. Thus:

$$L \geq L_1 + L_2 = A_1 + A_2 = A,$$

and therefore

$$\mathrm{mod}(B) = L^2/A \geq A = A_1 + A_2 = L_1^2/A_1 + L_2^2/A_2 = \mathrm{mod}(B_1) + \mathrm{mod}(B_2).$$

∎

The scaling of the metrics so $A_i = L_i$ is the same as scaling so the *circumferences* of these two annuli are the same; this is the right condition for optimal gluing. More generally, if $B_i \subset B$ are disjoint, essential annuli, then:

$$\mathrm{mod}(B) \geq \sum \mathrm{mod}(B_i).$$

It is not necessary that the B_i fill B. In particular, if $B' \subset B$, then

$$\mathrm{mod}(B') \leq \mathrm{mod}(B).$$

Similarly, when quadrilaterals are glued together in a suitable sense, we have

$$\mathrm{mod}(Q_1 \cup Q_2) \geq \mathrm{mod}(Q_1) + \mathrm{mod}(Q_2).$$

In both cases, there is no general reverse inequality.

Example: Bounds for a slit disk. Consider the annulus, defined for $R > 1$,

$$B(R) = \Delta(R) - [-1, 1].$$

We claim

$$\mathrm{mod}\, B(R) = (\log R)/(2\pi) + O(1). \tag{6.3}$$

66

In other words, mod $B(R)$ is within an additive constant of mod $A(R)$.

For the lower bound, just observe that $A(R)$ is contained in $B(R)$. What about an upper bound? Here are two methods:

Metrics. Define a conformal metric on $B(R)$ so ρ is 10 times the Euclidean metric on $|z| < 2$ and $\rho = |dz|/|z|$ on $|z| \geq 2$. Then

$$A(B(R), \rho) = 4\pi + 2\pi \log(R/2) = 2\pi \log R + O(1).$$

Now consider an essential loop $\gamma \in \Gamma_\theta(B(R))$. We distinguish three cases.

(i) If γ meets the circles $|z| = 1$ and $|z| = 2$, then $L(\gamma, \rho) \geq 10$, this being 10 times the Euclidean length of any arc joining these circles.

(ii) If γ meets the circle $|z| = 1$ but does not meet $|z| = 2$, then $L(\gamma, \rho) \geq 20$, since an loop enclosing $[-1, 1]$ has Euclidean length at least 2.

(iii) If γ does not meet the circle $|z| = 1$, then we use the fact that $\rho \leq |dz|/|z|$ on $B(R) - \Delta$ to conclude that $L(\gamma, \rho) \geq 2\pi$.

Putting this all together give $L(\Gamma_\theta, \rho) \geq 2\pi$, and thus

$$\mathrm{mod}(B(R)) \leq 1/\lambda(\Gamma_\theta) \leq A/L^2 = (\log R + 2)/(2\pi) = (\log R)/(2\pi) + O(1).$$

Note: if we had just used the Euclidean metric, the bound would be much worse — it would be off by a factor of $4\pi^2$, rather than by $O(1)$.

Conformal maps. Alternatively, consider the rational function $f(z) = (z + 1/z)/2$. This map sends S^1 to $[-1, 1]$. Let $B'(R) = f^{-1}(B(R))$, or more precisely the component of the inverse image lying outside the unit circle. Now it is easy to see that if $|f(z)| = R$, and $|z| > 1$, then $|z| < 2R + 1$, for R large. Thus $B'(R) \subset A(2R + 1)$, and we get

$$
\begin{aligned}
\mathrm{mod}(B(R)) &= \mathrm{mod}(B'(R)) \leq \mathrm{mod}(A(2R + 1)) = (\log R + \log 2 + 1)/(2\pi) \\
&= (\log R)/(2\pi) + O(1).
\end{aligned}
$$

Modulus and Dirichlet energy. In conclusion we remark that extremal length is also connected to potential theory. For example, let A be an annulus and let $\phi : A \to [0, 1]$ be the solution to the Dirichlet problem with boundary values 0 and 1 on the two components of ∂A. We then have:

$$1/\mathrm{mod}(A) = D(\phi) = \int_A |\nabla \phi|^2.$$

Note also that ϕ minimizes the Dirichlet energy among all functions with the given boundary values; thus any particular solution to the boundary value problem gives a lower bound, $\mathrm{mod}(A) \geq 1/D(\psi)$.

In the case of a Euclidean cylinder A of height h and circumference c, the desired function is $\phi(x,y) = y/h$; we have $|\nabla\phi|^2 = 1/h^2$, $\text{area}(A) = ch$ and hence the Dirichlet energy is $c/h = 1/\operatorname{mod}(A)$.

A similar statement holds for quadrilaterals. On $R(a,b)$, the harmonic function ϕ which is 0 on one of the sides of length a, and 1 on the other, has gradient $1/b$; hence

$$\int |\nabla\phi|^2 = a/b = \operatorname{mod} R(a,b).$$

Modulus and resistance. The modulus of an annulus or a quadrilateral can also be interpreted as its *resistance* in an electrical circuit. In such as circuit the voltage change, current flow and resistance are related by $V = IR$. For $V = 1$, the current flow I is given in terms of the solution to the Dirichlet problem above by the flux of $\nabla\phi$; for a cylinder of height h and circumference c, this is $I = c/h$, and hence $R = h/c = \operatorname{mod}(A)$.

7 Random walks on a lattice

In this section we begin the study random walks on the lattice $\mathbb{Z}^d \subset \mathbb{R}^d$ and their relation to harmonic functions. Our goals are:

1. To show the close relationship between random processes and harmonic functions in the discrete setting;

2. To show that discrete harmonic functions on $\mathbb{Z}^2$ converge, in the continuum limit, to smooth ones on domains in $\mathbb{C}$; and

3. To show that random walks converge, in the continuum limit, to Brownian motion.

The last statement allows stochastic methods to be applied directly to complex analysis, and leads to natural, probabilistic interpretations of harmonic measure, capacity and other analytic constructions.

A good resource here is [DY]; for general results in probability theory, see e.g. [Fe].

Background. The formal setup for probability theory is an abstract measure space $(\Omega, \mathcal{A}, \mu)$ consisting of a set Ω, a σ–algebra $\mathcal{A}$ of measurable sets, and a countably–additive probability measure

$$\mu : \mathcal{A} \to [0,1].$$

An *event* is a measurable set $A \in \mathcal{A}$; its *probability* is $\mu(A)$. A (real–valued) *random variable* is a measurable function

$$X : \Omega \to \mathbb{R};$$

its *expectation* is given by

$$E(X) = \int_\Omega X(\omega)\, \mu,$$

provided this integral exists. For any Borel set $B \subset \mathbb{R}$, a random variable X determines the event $(X \in B)$; the collection of all events determined by X gives a σ–subalgebra $\mathcal{A}_X \subset \mathcal{A}$.

Events A_1 and A_2 are *independent* if

$$P(A_1 \cap A_2) = P(A_1)P(A_2); \tag{7.1}$$

equivalently, if the conditional probability

$$P(A_1|A_2) = \frac{P(A_1 \cap A_2)}{P(A_2)} = P(A_1).$$

Note that independence is preserving under replacing an event by its negation, $A_i' = \Omega - A_i$.

Similarly, random variables X_1 and X_2 are independent if any two events they determine are independent; that is, equation (7.1) holds for all $A_i \in \mathcal{A}_{X_i}$, $i = 1, 2$. Independent random variables satisfy

$$E(X_1 X_2) = E(X_1)E(X_2).$$

There are similar definitions for collections of events. For example, $A_1, A_2, \ldots$ are independent if for all n, and all events B in the σ–algebra generated by $(A_i\ :\ i \neq n)$, we have

$$P(A_n|B) = P(A_n).$$

Mean, variance and standard deviation. The *mean* of a random variable X is given by $m = E(X)$. Its *variance* is given by

$$\mathrm{Var}(X) = E((X - m)^2);$$

the square–root of the variance is its *standard deviation* σ. It is readily verified that if X_1 and X_2 are independent, then

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2). \tag{7.2}$$

Thus the standard deviation σ of $X = X_1 + X_2$ satisfies the Pythagorean rule:
$$\sigma^2 = \sigma_1^2 + \sigma_2^2.$$

Not all random variables have a well–defined mean or variance. For example, if θ is chosen at random in S^1, then $E(|\tan\theta|) = \infty$. Thus $E(\tan\theta)$ is undefined (although one could argue that, by symmetry, it should be zero.)

Borel–Cantelli lemma. Given a sequence of events $(A_1, A_2, \ldots)$, let
$$B = \limsup A_i = \bigcap_i \bigcup_{j>i} A_j.$$

The event B occurs if and only if infinitely many A_i occur. We then have:

Lemma 7.1 *If $\sum P(A_i) < \infty$, then $P(\limsup A_i) = 0$. If $\sum P(A_i) = \infty$ and the events A_i are independent, then $P(\limsup A_i) = 1$.*

Proof. The first statement is immediate; for the second, observe that if $\sum P(A_i) = \infty$, then the probability that A_n does *not* occur for all $n > N$ is given by
$$\prod_N^\infty 1 - P(A_i) = 0.$$
∎

Kolmogorov's 0–1 law. Let X_i be a sequence of *independent* random variables. A *tail event* A is an event that is defined in terms of the variables X_i, but which does not depend on any particular variable. For example, the event
$$\limsup X_i > 1$$
is a tail event.

Theorem 7.2 *Any tail event has $P(A) = 0$ or $P(A) = 1$.*

Proof. The event A can be approximated by events A_n that depend only on $X_1, \ldots, X_n$, in the sense that $P(A \triangle A_n) \to 0$. By independence,
$$P(A) = \lim P(A \cap A_n) = \lim P(A)P(A_n) = P(A)^2.$$
Thus $P(A) = 0$ or $P(A) = 1$. ∎

70

The approximation result quoted in the proof is similar to the following result for Lebesgue measure: any measurable set $A \subset [0, 1]$ can be approximated by a finite union of intervals B, in the sense that $m(A \triangle B) < \epsilon$.

The Fourier transform. We will take advantage of the fact that $G = \mathbb{Z}^d$ is a *discrete Abelian group* in our study of random walks. Its *dual group* is given by

$$\widehat{G} = \mathrm{Hom}(G, S^1) = (S^1)^d.$$

Here S^1 can be thought of as either the unit circle $|z| = 1$ in $\mathbb{C}$, or as the group $\mathbb{R}/2\pi\mathbb{Z}$ with parameter θ. The character

$$\chi : \mathbb{Z}^d \to S^1$$

in $\widehat{G}$ determined by $\theta = (\theta_1, \ldots, \theta_d) \in (S^1)^d$ is given by:

$$\chi(x) = \exp(ix \cdot \theta) = \prod_{i=1}^{d} \exp(ix_i \theta_i).$$

The dual of $\widehat{G}$ is again G. Each group carries a natural measure – counting measure on $\mathbb{Z}^d$, and Haar measure on $(S^1)^d$ (with total mass one).

The *Fourier transform* is a natural isomorphism between $L^2(G)$ and $L^2(\widehat{G})$. In this case, the Fourier transform of

$$f : \mathbb{Z}^d \to \mathbb{C}$$

is the function $\widehat{f} : \widehat{G} \to \mathbb{C}$ given by

$$\widehat{f}(\chi) = \int_G f(x)\overline{\chi}(x)\,dx,$$

or more concretely by

$$\widehat{f}(\theta) = \sum_{x \in \mathbb{Z}^d} f(x)\exp(-ix\theta).$$

Provided it is well–behaved, the original function $f(x)$ can be recovered from its Fourier transform by:

$$f(x) = (2\pi)^{-d}\int \widehat{f}(\theta)\exp(i\theta)\,d\theta.$$

71

Thus the Fourier transform describes how to decompose $f(x)$ as a (continuous) linear combination of characters on $\mathbb{Z}^d$.

In probability theory, a random variable X with values in $\mathbb{Z}^d$ is described by a function $p(x) \geq 0$ on $\mathbb{Z}^d$ with $\sum p(x) = 1$. The Fourier transform of p is called the *characteristic function* of X; it is given by

$$F_X(\theta) = \widehat{p}(-\theta) = E(\exp(i\theta X)).$$

Since $p \in L^1(G)$, the function $\widehat{p}(\theta)$ is continuous.

Passage to F_X simplifies the study of sums of random variables; we have:

Proposition 7.3 *If* $Y = X_1 + \cdots + X_n$ *is a sum of independent random variables, then*

$$F_Y(\theta) = \prod_1^n F_{X_i}(\theta).$$

Generating functions. For a random variable τ taking values in $\mathbb{Z}$, the Fourier transform can alternatively be described as the *generating function* $F_\tau(z)$, where $z = \exp(i\theta)$. It is given by

$$F_\tau(z) = E(z^\tau) = \sum P(\xi = n)z^n.$$

If τ and σ are independent, then

$$F_{\tau+\sigma}(z) = F_\tau(z)F_\sigma(z).$$

For example, if $\tau = \pm 1$ with equal probabilities, then

$$F_\tau(z) = (z + z^{-1})/2.$$

Example: a wide variety of coins. Suppose we have N coins that come up heads with probabilities $p_1, \ldots, p_N$. How can one efficiently compute the probability P_k that exactly k come up heads?

Answer: we have

$$\sum P_k z^k = \prod_1^N (q_i + p_i z),$$

where $p_i + q_i = 1$, and this product can be quickly computed. In fact if $\tau = 1$ with probability p and 0 with probability $q = 1 - p$, then $F_\tau(z) = q + pz$. So this is just an application of the method of generating functions.

Note that if we are only interested in P_k with k small, we can compute this product even more rapidly, since only powers z^i with $i \leq k$ need to be retained while forming the product.

This application has practical importance in biological statistics. E.g. one might be testing a vaccine on a large population of various ages, where the probability of infection varies with age. Then the efficacy of a vaccine involves comparing the expected infection rate to the observed one.

The Gaussian distribution. The standard Gaussian distribution on $\mathbb{R}$ is given by the density

$$G(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) :$$

it satisfies

$$\int G(x)\, dx = \int x^2 G(x) = 1.$$

Thus $G(x)$ describes the distribution of a Gaussian random variable with mean $E(X) = 0$ and variance $E(X^2) = 1$. More generally,

$$G_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-x^2/(2\sigma^2))$$

describes a Gaussian random variable X_σ with mean zero and standard deviation σ. The sum of two independent Gaussians is again a Gaussian, and the variances add; informally, we have:

$$X_{\sigma^2} + X_{\delta^2} = X_{\sigma^2+\delta^2}.$$

Stationary phase. Here is a general result that explains the appearance of the Gaussian distribution in many settings.

Let $f : [-b, b] \to [0, 1]$ be a smooth function with a $f(0) = 1$, $a = f''(0) < 0$, and $f(x) < 1$ for $x \neq 0$. Recall that $(1 - x/n)^n \approx \exp(-x)$. We then have

$$\begin{aligned} f(x)^n &\approx (1 - ax^2/2)^n = (1 - anx^2/(2n))^n \approx \exp(-nax^2/2) \\ &= \exp(-x^2/(2\sigma^2)), \end{aligned}$$

where $\sigma = 1/\sqrt{an}$. (As a quick check, compare the coefficients of x^2 in the power series at $x = 0$ on both sides.) It follows that

$$\int_{-a}^{a} f(x)^n\, dx \sim \sigma\sqrt{2\pi} = \sqrt{2\pi}/\sqrt{an}.$$

For example, when n is large and even, the average value of $\cos^n(x)$ is given by

$$C_n = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \cos^n(x)\, dx \sim \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{n}}. \tag{7.3}$$

(It is a classical problem of Arnold to estimate this average by hand for $n = 100$, with an error of at most 20%. Using $\sqrt{50\pi} \approx 7 \cdot 1.7 \approx 12$ gives $C_n \approx 1/12 \approx 0.08333\ldots$, versus the exact value $0.079589\ldots$)

Random walks. Let $\xi_1, \xi_2, \ldots$ be a sequence of independent random variables, each assuming the values ± 1 with equal probability. The symmetric random walk on $\mathbb{Z}$ is the sequence of random variables

$$x(n) = \sum_1^n \xi_i.$$

Here the random walks starts at $x(0) = 0$.

The symmetric random walk on $\mathbb{Z}^d$ is defined similarly: if (e_j) form a standard basis for $\mathbb{R}^d$, then ξ_i assumes the $2d$ values $(\pm e_j)$ with equal probability.

We can also shift the basepoint and specify that $x(0) = x \in \mathbb{Z}^d$; then

$$x(n) = x(0) + \sum_1^n \xi_i.$$

Our intuition about a random walk is guided by the following basic result:

Theorem 7.4 *A random walk in $\mathbb{Z}^d$ based at $x(0) = 0$ satisfies*

$$E(\|x(n)\|^2) = n.$$

Thus we should think of random walker as moving rather slowly with respect to time, and staying (with high probability) within a ball of radius $O(\sqrt{n})$ up to time n.

Recurrence in $\mathbb{R}^d$. In dimension d, there are $(2d)^n$ possible random walks of length d starting at the origin. How many of these return to the origin? We will see:

Theorem 7.5 *In dimension d, the probability $r_d(2n)$ that a path starting at 0 returns to 0 at time $2n$ is asymptotic to a multiple of $n^{-d/2}$.*

A walk $x(n)$ on $\mathbb{Z}^d$ is *recurrent* if it returns infinitely often to a bounded set. Recurrence of a random walk is a *tail event*, and thus it has probability 0 or 1.

It is easy to see that a recurrent random walk almost surely visits *every* point in $\mathbb{Z}^d$ infinitely often. Thus to test recurrence, it suffices to test for infinitely many returns to $x = 0$. With this in mind, the result above implies:

Corollary 7.6 *A random walk is recurrent in $\mathbb{Z}^d$ when $d = 1$ or 2, while $x(n) \to \infty$ with probability one when $d \geq 3$.*

Proof. For $d \geq 3$, $\sum r_d(n) < \infty$ and hence with probability one $x(n) = 0$ for only finitely many n. It follows that for each fixed y, $x(n) = y$ for only finitely many n, since after each visit to y there is a definite probability of a visit to 0. Thus $x(n) \to \infty$.

For $d \leq 2$, we have $R = \sum r_d(n) = \infty$. This sum R is the expected total number of visits to $x = 0$. We also have $R = \sum_1^\infty r^n$, where r is the probability that $x(n) = 0$ for some $n > 0$. Thus $r = 1$ and hence $x(n)$ returns to zero at least once with probability one. Once it has done so, the probability of returning again remains one. Thus $x(n)$ is recurrent. ∎

The mechanism behind Theorem 7.5 is easy to describe: $x(n)$ is fairly evenly distributed in the ball of radius $\sqrt{n}$, and the number of lattice point in this ball is comparable to $n^{d/2}$, so the probability that $x(n) = 0$ is comparable to $n^{-d/2}$.

Note however that by parity considerations, a path must have even length to return to the origin.

Dimension 1. We now turn to a more precise discussion. A path of length $2n$ returning to the origin must consist of n positive steps and n negative steps. Thus, the number of paths returning to 0 of length $2n$ is

$$R_1(2n) = \binom{2n}{n}.$$

Alternatively, the probability of returning to the origin is

$$r_1(n) = \frac{1}{2\pi} \int_0^{2\pi} \cos^n(\theta)\, d\theta. \tag{7.4}$$

We have $r_1(n) = 2^{-n}R_1(n)$. Using Stirling's formula in the first case, and equation (7.3) in the second, we find:

$$r_1(2n) \sim \frac{1}{\sqrt{\pi n}}.$$

Here are two explanations of formula (7.4). Let $p_n(i) = P(x(n) = i)$ describe the distribution of the random walk on $\mathbb{Z}$ after n steps. It is then easy to see that

$$\left(\frac{z + z^{-1}}{2}\right)^n = \sum_i p_n(i)z^i.$$

Thus $r_1(n) = p_n(0)$ is the constant coefficient above, and it can be found by setting $z = \exp(i\theta)$; we then have

$$\left(\frac{z + z^{-1}}{2}\right)^n = \cos^n \theta,$$

and averaging over S^1 gives the constant term.

Equivalently, if ξ assumes the values ± 1 with equal probability, then

$$F_\xi(\theta) = E(e^{i\xi\theta}) = \cos(\theta).$$

Since $x(n) = \xi_1 + \cdots + \xi_n$, we have

$$F_{x(n)}(\theta) = F_\xi(\theta)^n = \cos^n(\theta).$$

Inverting the Fourier transform, we find that for $d = 1$ the probability of a path from 0 to x is given by:

$$p_n(0, x) = \frac{1}{2\pi} \int_{S^1} \exp(ix\theta) \cos^n(\theta) \, d\theta;$$

and setting $x = 0$ gives equation (7.4).

Dimension 2. If $x(n) = (a(n), b(n))$ is a random walk in $\mathbb{Z}^2$, then $a(n)+b(n)$ and $a(n) - b(n)$ are independent random walks in $\mathbb{Z}$. It follows that

$$R_2(n) = R_1(n)^2 \quad \text{and} \quad r_2(n) = r_1(n)^2;$$

in particular we have

$$r_2(2n) \sim \frac{1}{\pi n}.$$

Unfortunately this type of simplification does not occur in higher dimensions.

Dimension d. For a random walk on $\mathbb{Z}^d$, we have

$$F_\xi(\theta) = \frac{1}{d} \sum_1^d \cos(\theta_i),$$

and the Fourier method yields:

$$r_d(n) = \frac{1}{(2\pi)^d} \int_{(S^1)^d} \left(\frac{1}{d} \sum \cos(\theta_i) \right)^n d\theta. \tag{7.5}$$

One also has a combinatorial formula:

$$R_d(n) = \sum_{n_1 + \cdots + n_d = n} \frac{n!}{n_1! n_2! \cdots n_d!} R_1(n_1) R_1(n_2) \cdots R_1(n_d),$$

where $n_i \geq 0$ counts the number of steps along the ith axis. We have $r_d(n) = (2d)^{-n} R_d(n)$.

Proof of Theorem 7.5. We will be content to show that $r_d(n) \asymp n^{-d/2}$. We take equation (7.5) as our point of departure; it says that

$$r_d(n) = (2\pi)^{-d} \int_{(S^1)^d} \Phi(\theta)^n \, d\theta,$$

where

$$\Phi(\theta) = \frac{1}{d} \sum_1^d \cos(\theta_i).$$

The main contribution to the integral occurs near the finitely many points where $\Phi(\theta) = \pm 1$. We will examine the contribution near $\theta = 0$; the behavior at the other points is essentially the same, provided n is even. For θ near zero, we have

$$\Phi(\theta) \approx 1 - (1/d) \sum \theta_i^2 / 2 = 1 - r^2/(2d),$$

where $r^2 = \sum \theta_i^2$. Thus $\Phi(\theta)^n$ is approximately a Gaussian with height 1 and $\sigma \asymp 1/\sqrt{n}$. Thus its integral is comparable σ^d, the volume of a ball of radius σ in $\mathbb{R}^d$. This gives $r_d(2n) \asymp \sigma^d \asymp n^{-d/2}$. $\blacksquare$

Discrete harmonic functions. Let us define an averaging operator for functions $f : \mathbb{Z}^d \to \mathbb{R}$ as follows:

$$Pf(x) = \frac{1}{2d} \sum_{|y-x|=1} f(y).$$

The sum is over the $2d$ lattice points y that are adjacent to x, in the sense the max $|x_i - y_i| = 1$.

In terms of this operator, an analogue of the Laplacian is defined by

$$Af(x) = Pf(x) - f(x).$$

In dimension $d = 1$, we have

$$Af(x) = (1/2)(f(x+1) + f(x-1) - 2f(x)),$$

which is an approximation to $(1/2)f''(x)$. (In fact if $f(x) = ax^2 + bx + c$, then $Af(x) = a = (1/2)f''(x)$.)

In general, Af plays the role of $(1/2d)\Delta f$ on $\mathbb{R}^d$. We say f is *harmonic* if $Af = 0$, or equivalently if $Pf = f$.

Dirichlet energy. To justify this analogy, we also define the Dirichlet energy of f by

$$D(f) = \sum_x |\nabla f|^2(x) = \frac{1}{2d} \sum_{|x-y|=1} |f(x) - f(y)|^2.$$

Note that there are $2d$ oriented edges (x, y) per vertex, in the adjacency graph of $\mathbb{Z}^d$. Thus in the sum on the right, each oriented edge is weighted to achieve a density of one. One could also sum over unoriented edges, each with weight $1/d$.

Proposition 7.7 *If $f : \mathbb{Z}^d \to \mathbb{R}$ has finite support, then*

$$\sum |\nabla f|^2(x) = -2 \sum f(x)Af(x).$$

Proof. We have

$$
\begin{aligned}
\frac{1}{2}D(f) &= \frac{1}{4d} \sum_{|x-y|=1} (f(x) - f(y))^2 = \frac{1}{4d} \sum_{|x-y|=1} f(x)^2 + f(y)^2 - 2f(x)f(y) \\
&= \sum_x f(x)^2 - \frac{1}{2d} \sum_{|x-y|=1} f(x)f(y) \\
&= \sum_x f(x) \left(f(x) - \frac{1}{2d} \sum_{|x-y|=1} f(y) \right) = -\sum f(x)Af(x).
\end{aligned}
$$

78

Here we have used the fact that there are $2d$ values of y with $|x - y| = 1$. ∎

The maximum principle. Many results for harmonic functions on $\mathbb{R}^d$ carry over to $\mathbb{Z}^d$. To state on such result, let $V \subset \mathbb{Z}^d$ be a finite set. The *boundary* of V is defined by

$$\partial V = B = B(V, 1) - V. \qquad (7.6)$$

A point $y \in \mathbb{Z}^d$ belongs to ∂V iff $y \notin V$ but $|x - y| = 1$ for some $x \in V$. We say V is *connected* if any two points can be joined by a path x_i with $|x_{i+1} - x_i| = 1$.

Theorem 7.8 *Suppose $f : V \cup \partial V \to \mathbb{R}$ is harmonic on V. Then*

$$\max_V f(x) = \max_{\partial V} f(x).$$

Moreover, if V is connected and f achieves its boundary at an interior point, then f is constant.

Proof. If f achieves its maximum M at $x \in V$, then $f(y) = M$ for any y with $|x - y| = 1$. Following a chain of consecutive y's leads to the boundary, and also connects x to any other point of V provided V is connected. ∎

Bounded harmonic functions. Although the next result is familiar from $\mathbb{R}^d$, the proof is rather different.

Theorem 7.9 *A bounded harmonic function f on $\mathbb{Z}^d$ is constant.*

Proof. Suppose f achieves its maximum M at x. Since $Pf(x) = M$, we must have $f(y) = f(x)$ for all points with $|x - y| = 1$. It quickly follows that $f(x) = M$ for all x.

Now suppose f does not achieve its maximum M. Then $\min f < \max f$; changing sign, we can assume $M > 0$. Now whenever $f(x)$ is very close to M, and $|x - y| = 1$, $f(y)$ must also be very close to M. Thus there are arbitrarily large balls on which $f(x)$ is close to M. In particular, for each n we can find x such that

$$\sum_{k=0}^{n} f(x + ke_1) \geq nM \gg 0.$$

79

It follows that f is not the *coboundary* of a bounded function $h(x)$; that is, we cannot write

$$f(x) = h(x + e_1) - h(x)$$

for some bounded function h. If we could, the sums above would telescope and be bounded by $2 \max |h(x)|$.

Now for any bounded harmonic function $F(x)$, the coboundary (or discrete derivative)

$$f(x) = F(x + e_1) - F(x)$$

is also bounded and harmonic. By the reasoning above, $f(x)$ achieves its maximum, and hence it is constant. Since F is bounded, this constant must be zero. The same holds with e_1 replaced by e_i, $i = 1, 2, \ldots, d$ and thus $F(x)$ is constant. $\blacksquare$

Random walks and harmonic functions. We now fix the dimension d. For a random walk on $\mathbb{Z}^d$, we let $p_n(x, y)$ be the probability that a random walk starting at x reaches $y \in \mathbb{Z}^d$ at its nth step.

The fundamental link between random walks and harmonic functions is provided by the following simple observation: we have

$$E_x(f(x(1))) = E(f(x + \xi_1)) = Pf(x).$$

More generally, we have:

$$E_x f(x(n)) = (P^n f)(x) = \sum_y p_n(x, y) f(y).$$

Martingales. In probability theory, one says that a sequence of random variables $X_1, X_2, \ldots$ form a *martingale* if

$$E(X_{n+1} | X_1, \ldots, X_n) = X_n.$$

For example, if a gambler plays a sequence of independent fair games of chance, and after the nth game his holdings total X_n, then X_n forms a martingale. The observation above can be stated as follows:

Theorem 7.10 *Let $x(n)$ be a random walk on $\mathbb{Z}^d$ starting at $x(0) = x$. Then $f : \mathbb{Z}^d \to \mathbb{R}$ is harmonic if and only if the random variables $f(x(n))$ form a martingale.*

Note: superharmonic functions, satisfying $f \geq P(f)$, correspond to supermartingales, or losing games, where $E(X_{n+1}|\ldots) \leq X_n$.

The heat equation. The heat equation on $\mathbb{R}^n$ is given by

$$\frac{df_t}{dt} = \Delta f_t.$$

It is a parabolic differential equation that can be solved for $t \geq 0$ with suitable initial data $f_0 : \mathbb{R}^n \to \mathbb{R}$ (e.g. a compactly supported smooth function). Note that over time, f_t increases at its minima and decreases at its maxima; in fact f_t gets smoother as $t \to \infty$. It is also true that $\int f_t$ is constant in time, since $\int \Delta f_t = 0$.

A discrete analogue of the heat equation, on $\mathbb{Z}^d$, is given by

$$f_{n+1} - f_n = A f_n. \tag{7.7}$$

It is defined for time $n \geq 0$ once we have fixed $f_0 : \mathbb{Z}^d \to \mathbb{R}$. Since $Af = Pf - f$, its solution is given by

$$f_n(x) = P^n f_0(x) = \sum_y p_n(x,y) f_0(y). \tag{7.8}$$

The quantity $p_n(x,y)$ is therefore called the *heat kernel*. The fact that $\sum_y p_n(x,y) = 1$ shows that $\sum_x f_n(x)$ is independent of n.

The Green's function for $\mathbb{Z}^d$. Next we aim to solve the inhomogeneous Laplace equation

$$Af = -h$$

for a given compactly supported h on $\mathbb{Z}^d$.

The solution we give we will work when $d \geq 3$. First, we define the *Green's function* for $\mathbb{Z}^d$ by

$$G(x,y) = \sum_{n=0}^{\infty} p_n(x,y).$$

Note that $G(x,y) = G(y,x) = G(x-y,0)$. The Green's function has a simple interpretation:

> $G(x,y)$ *is the expected number of visits to y made by a random walk with $x(0) = x$.*

Recall that for $d \geq 3$, we have

$$G(0,0) = \sum_n r_d(n) \asymp \sum_n n^{-d/2} < \infty,$$

and similar reasoning shows $G(x,y)$ is finite for all $d \geq 3$. It can also be shown that for fixed x, $G(x,y) \to 0$ as $y \to \infty$.

The corresponding operator G acts on functions by

$$Gf(x) = \sum_y G(x,y)f(y).$$

In terms of a solution to the heat equation with $f_0 = f$, we have

$$Gf = \sum_0^\infty f_n = \sum_0^\infty P^n(f).$$

In view of (7.7), if we apply A to this sum it telescopes, giving

$$AGf = -f.$$

(Formally, we have just used the fact that $(I - P)^{-1} = \sum P^n$.) This shows:

Theorem 7.11 *For $d \geq 3$, and any compactly supported function h, $f = Gh$ is the unique solution to $Af = -h$ which vanishes at infinity.*

The smooth Green's function. We remark that for a compactly supported smooth function h on $\mathbb{R}^d$, the solution to $\Delta f = -h$ is given by convolution with

$$G(x) = C_d/|x|^{d-2},$$

where C_d is chosen so the flux of G through the unit sphere is -1, i.e. $\Delta G = -\delta_0$. Note that $G(x)$ vanishes at infinity.

Visiting neighbors. One can also see directly that $AG(x,0) = -\delta_0$.

By definition, $G(x,0)$ is the expected number of visits that a random walk, starting at 0, makes to x. Provided $x \neq 0$, every visit to x must be preceded by a visit to one of its neighbors. Thus $G(x,0)$ is the average of the number of visits to its neighbors, and hence $G(x,0)$ is harmonic at x.

The same reasoning applies when $x = 0$, except there is one occasion when the path can visit 0 without visiting a neighbor first: when $x(0) = 0$. Thus

$$G(x,0) = \delta_0 + PG(x,0),$$

and hence $AG = PG - G = -\delta_0$.

Payoffs. The equation

$$f(x) = Gh(x) = \sum_n p_n(x, y)h(y)$$

leads to the following interpretation of f. Suppose the random walker collects a payoff of $h(y)$ whenever they visit y. Then $f(x)$ is the total expected payoff, along an infinite walk starting at x.

The case $d \leq 2$: renormalization. One can also solve Laplace's equation for $d \leq 2$, but the Green's function must be defined differently. Namely we adjust it by a constant, as n varies, so it converges. Thus one defines:

$$G(x, y) = \sum_n p_n(x, y) - p_n(0, 0).$$

The point is that $p_n(x, y) = p_n(0, x - y)$ grows, with n, at about the same rate as $p_n(0, 0)$. Subtracting off this constant leaves us with a convergent series. One can then show that $A_x G(x, y) = -\delta_{xy}$. But in this case $G(x, y)$ has a logarithmic pole at infinity.

Solving the Dirichlet problem. We now turn to harmonic functions on bounded domains. This part of the discussion works for any dimension d.

The Dirichlet problem in the discrete setting can be formulated as follows. Let $V \subset \mathbb{Z}^d$ be a finite set. Given

$$\Phi : \partial V \to \mathbb{R},$$

we wish to extend Φ to a function

$$\phi : V \cup \partial V \to \mathbb{R}$$

such that ϕ is harmonic on V. This function is unique by the maximum principle.

The solution is easily described in terms of random walks. First, note that a random walk $x(n)$ initiated at $x \in V$ will eventually hit B, with probability one. This is easy to prove directly; it also follows from Kolmogorov's 0/1–law, since the probability is a tail event and it is clearly positive.

We wish to modify the random walk so that once it reaches B, it stops. A convent way to make this modification is do define the *stopping time τ* to be the least n such that $x(n) \in B$. Here τ is a random variable that depends on the entire walk, including its starting point. If $x = x(0) \in B$, then $\tau = 0$.

Theorem 7.12 *The solution to the Dirichlet problem on V is given by*

$$\phi(x) = E_x(\Phi(x(\tau))).$$

Proof. If $x \in V$ then $\tau \geq 1$. Thus the random walk from x to B can be constructed by first choosing a neighboring vertex y at random, and then taking a walk starting at y. This shows ϕ is harmonic at x; and $\phi|B = \Phi$ since there $\tau = 0$. ∎

Note: Existence proof. It is not hard to prove the Dirichlet problem *has* a solution by counting the dimensions of the vector spaces involved. The point of using random walks is to gain insight into the meaning and behavior of the solution.

Harmonic measure. For example, there is a natural and elegant interpretation of the *harmonic measure* of a set $A \subset \partial V$. As before, the *definition* of this measure is

$$\omega(x, V, A) = \phi(x),$$

where ϕ is harmonic on V and $\phi|\partial V = \chi_A$. We then have:

Proposition 7.13 *The harmonic measure of $A \subset \partial V$ as seen from $x \in V$ is the probability that a random walk, starting at x, lands in A when it first exits V.*

Note the similarity to the unit disk, when $\omega(z, \Delta, A)$ is the probability that a random *hyperbolic geodesic ray* based at z lands on $A \subset S^1$.

Example: escape from an interval. What is the probability $p(x)$ that a random walk starting at $x(0) = x$ in $[0, N]$ hits N before hitting 0? The answer is given by the harmonic function

$$p(x) = x/N.$$

This is the harmonic measure of the endpoint N as seen from x. Note that it is easier to determine $p(x)$ for all x at once, than to try to compute it for an individual value of x, using random walks!

Gambler's ruin revisited. Suppose we have \$1000 to gamble with at a roulette table. We bet \$1 on red every time, and quit once have made \$50. What are the chances of success?

Let us first assume our winning chances on each spin are $p = 1/2$. Then we are taking a random walk on $[0, 1050]$. The probability that, starting with x dollars, our holdings reach \$1050 before we run out of money, is given by the harmonic function $s(x) = x/1050$. We conclude that:

The chances of winning \$50 are $20/21 = 95.2\%$.

This is an example of a *fair game*; on average, our fortune is neither increased nor decreased, because:

Our expected gain is $50(20/21) - 1000/21 = 0$.

In fact, we could have used this fairness to give a direct calculation of the probability of success.

Expected length of the game. How long will our gambler take to finally win or be ruined? Let $f(x)$ be the expected length of the game (number of plays) starting with x dollars. Then $f(x) = 0$, $f(1050) = 0$, and

$$f(x) = 1 + (f(x-1) + f(x+1))/2. \tag{7.9}$$

Thus $f(x)$ is a solution to the *inhomogeneous equation $Af(x) = 1$*. The solution, with the given boundary conditions, is given by

$$f(x) = x(1050 - x). \tag{7.10}$$

(To check this, observe that $P(x^2) = x^2 + 1$.) In particular:

An average of 50,000 spins are required to complete the game, starting with \$1000.

Equally striking: if we start with \$1 and quit only if we make \$1000 or are ruined, then the average number of spins to complete the game is 1000. If the house has infinite resources, the expected waiting time for ruin is infinite.

The House's take. However, this is not how roulette works. In European roulette there is a green slot marked 0 that pays out to the house. Thus the probability of winning a bet on red is $p = 18/37 < 1/2$, and the probability of losing is $q = 19/37$. We now have:

$$s(x) = qs(x-1) + ps(x+1). \tag{7.11}$$

This is a *biased* random walk. Observe that one solution is given by

$$s(x) = (q/p)^x,$$

since
$$q(q/p)^{x-1} + p(q/p)^{x+1} = (q/p)^x (p + q) = (q/p)^x.$$

Another solution is given by $s(x) = 1$. To satisfy the boundary conditions $s(0) = 0$, $s(1050) = 1$, we take the linear combination

$$s(x) = \frac{(q/p)^x - 1}{(q/p)^{1050} - 1}$$

Thus the chances of winning \$50 are now:

$$s(1000) \approx (p/q)^{50} = (18/19)^{50} = 6.7\%.$$

Now our expected losses are not \$0, but rather about \$933.

Even though it seems like a small modification of a fair game, our strategy works very poorly when playing this modification. It is much better in these circumstances to make one large bet than many small ones. Betting all \$1000 at once gives an expected loss of $\$1000(q - p) \approx \$27$.

Note: in American roulette there are two green squares, and our chances of winning become $(18/20)^{50} = 0.5\%$. Our expected losses are about \$994.

The Green's function in a bounded domain. Let $V \subset \mathbb{Z}^d$ be a finite set, with boundary $B = \partial V$. Given $h : V \to \mathbb{R}$, we would like to solve the inhomogeneous equation
$$Af = -h,$$
subject to the condition $f|B = 0$.

In this case a suitably modified Green's function makes sense for all values of $d \geq 1$. To define it, we consider random walks $x(n)$ as usual and for $x, y \in V \cup B$ set

$$p_n^B(x, y) = P_x(x(n) = y \quad \text{and} \quad x(i) \notin B, i = 0, 1, 2, \ldots, n\}. \qquad (7.12)$$

This quantity is $(2d)^{-n}$ times the number of paths of length n from x to y that lie entirely in V. Clearly this count is symmetric in x and y; and we set

$$G(x, y) = \sum_{n=0}^{\infty} p_n^B(x, y).$$

Just as before, it is now easy to check that $G|B = 0$, and

$$A_x G(x, y) = -\delta_{xy},$$

and thus the unique solution to $Af = -h$, vanishing on B, is given by

$$f(x) = \sum_y G(x,y)h(y).$$

The probabilistic interpretation here is:

> $f(x)$ is the expected total payoff along a random walk that starts at x, ends when $x(n) \in B$, and accrues a payoff of $h(y)$ each time $x(i) = y$.

Example. Let $V = \{1, 2, \ldots, n-1\} \subset \mathbb{Z}$. Then for each fixed y, $G(x,y)$ is a linear function on $[0,y]$ and $[y,n]$, vanishing at the boundary points $0, n$ and achieving its maximum

$$G(y,y) = 2y(n-y)/n$$

at the point y. For $x < y$ we have

$$G(x,y) = 2x(n-y)/n.$$

To compute the value for $x > y$, one can use the fact that $G(x,y) = G(y,x)$.

Note that $G(x,x) - 1$ is the expected number of times that a random walk starting at x returns to x, before hitting 0 or n. It is maximized at $x = n/2$, where $G(x,x) = n/2$.

Our calculation of the exit time $f(x)$ for gambler's ruin, equation (7.10), is also an instance of the solution of the inhomogeneous equation on an interval in $\mathbb{Z}$.

8 More on random walks on $\mathbb{Z}$

In this section we develop random walks on $\mathbb{Z}$ in more detail. These results will be used on the next section to show that discrete harmonic functions converge to smooth ones.

The central limit theorem: distribution in space. We consider a random walk on $\mathbb{Z}$ starting at $x = 0$. We have seen that $x(n)$ is a random variable with mean zero and variance $E(x(n)^2) = n$. The *central limit theorem* states that $x(n)$ is distributed like a bell curve with the same invariants. See Figure 7 for an example.
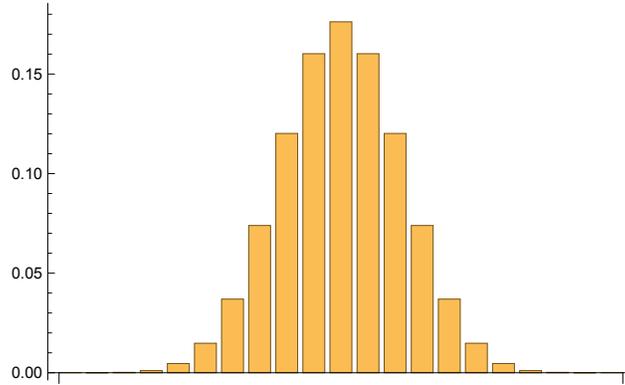
Figure 7. Distribution of a random walk on $\mathbb{Z}$ after 20 steps.

Theorem 8.1 *For all t we have*

$$P(x(n) > t\sqrt{n}) \to \frac{1}{\sqrt{2\pi}} \int_t^\infty \exp(-t^2/2)\, dt$$

as $n \to \infty$.

The resemblance between the continuous density $\exp(-t^2/2)$ and the discrete series $p_{2n}(0, 2i)$ is evident after taking logarithms of the second quantity; see the inverted parabola in Figure 8.

The Chernoff bound. This statement is slightly less than what we will need in applications, because it does not come with an error term. Instead we will use a weaker form that has the advantage of being both uniform in n and easy to prove. This is:

Theorem 8.2 *For any $k, n \geq 0$, with n an integer, we have*

$$P(x(n) \geq k\sqrt{n}) \leq \exp(-k^2/4).$$

The following bound by a geometric series follows immediately, and is useful to quote in applications:

Corollary 8.3 *Given $\Lambda > 1$, for all integers $k, n \geq 0$ we have*

$$P(|x(n)| \geq k\sqrt{n}) \leq C_\Lambda \Lambda^{-k}.$$
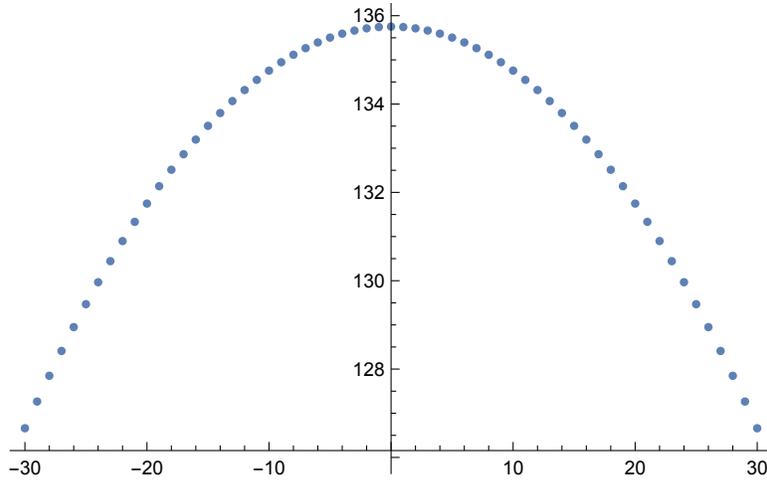
88

Figure 8. Plot of $\log p_{2n}(2i) = \log \binom{2n}{n+i}$, $n = 100$.

For example, this Corollary is more than enough to imply the *strong law of large numbers*:

Theorem 8.4 *We have $x(n)/n \to 0$ with probability one.*

Proof. Fix $r > 0$. Note that rn is on the order of $\sqrt{n}$ standard deviations away from 0. Thus Corollary 8.3 implies $P(|x(n)| > rn) = O(p(n))$, $p(n) = 2^{-\sqrt{n}}$). Since $\sum p(n) < \infty$, this event occurs for only finitely many n (Borel–Cantelli), and thus $x(n)/n \to 0$. ∎

Proof of Theorem 8.2. We first recall the elementary *Markov inequality*: if $X \geq 0$ is a non–negative random variable, then

$$P(X \geq a) \leq E(X)/a.$$

The clever step in the proof is to observe that for any $t \geq 0$, we have

$$P(x(n) \geq k\sqrt{n}) = P(\exp(tx(n)) \geq \exp(tk\sqrt{n})) \leq \frac{E(\exp(tx(n)))}{\exp(tk\sqrt{n})}.$$

We may assume $k \leq \sqrt{n}$ since $x(n) \leq n$. Let

$$t = \frac{k}{2\sqrt{n}} \leq \frac{1}{2}. \tag{8.1}$$

89

Next we observe that, since $x(n) = \sum \xi_i$ and $E(\exp(t\xi_i)) = \cosh(t)$, we have

$$E(\exp(tx(n))) = \cosh(t)^n \le (1 + t^2)^n \le \exp(nt^2).$$

Here we have used the general fact that $\log(1 + t) \le t$ for $t \ge 0$, and the more special fact that $\cosh(t) \le 1 + t^2$ for $t \le 1$. All together this gives

$$P(x(n) \ge k\sqrt{n}) \le \frac{\exp(nt^2)}{\exp(tk\sqrt{n})}.$$

The value of t in (8.1) has been chosen to minimize the expression above; the numerator becomes $\exp(k^2/4)$, the denominator becomes $\exp(k^2/2)$, and the desired bound follows. ∎

Ratios of binomial coefficients. We remark that for $p + q = 2n$, we have the exact formula

$$a_p = P(x(2n) = p - q) = 2^{-2n} \frac{(2n)!}{p!q!}.$$

It follows that

$$\frac{a_{p+1}}{a_p} = \frac{q}{p+1}.$$

Starting at $p = q = n$, one then gets

$$a_{n+k} = a_n \cdot \frac{n(n-1)\cdots(n-k+1)}{(n+1)(n+2)\cdots(n+k)}. \tag{8.2}$$

For k on the order of $\sqrt{n}$ or smaller, the product on the right is approximately

$$\prod_{j=1}^{k}(1 - j/n)^2 \approx \exp(-k^2/n),$$

since $2\sum_1^k j \approx k^2$. One can continue this analysis to obtain a proof of the central limit theorem for the distribution of $x(2n)$.

First return to zero: getting lost. Let $x(n)$ be a random walk on $\mathbb{Z}$ with $x(0) = 0$. We let

$$u_{2n} = 2^{-2n} \binom{2n}{n}$$

90

denote the probability that $x(2n) = 0$. As we have seen,

$$u_{2n} \sim \frac{1}{\sqrt{\pi n}}.$$

Closely related is the probability f_{2n} that the *first return* to the origin occurs when $x(2n) = 0$.

Theorem 8.5 *The probability that the first return to the origin occurs when $x(2n) = 0$ is given by*

$$f_{2n} = \frac{1}{2n-1} u_{2n}. \tag{8.3}$$

Corollary 8.6 *The expected waiting time for the first return to $x = 0$ is infinite.*

Proof. We have $\sum (2n) f_{2n} = \infty$. ∎
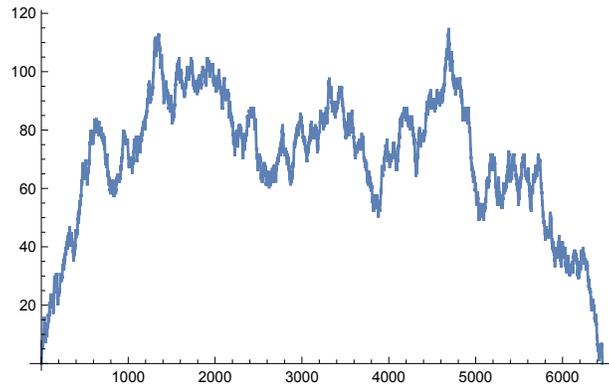
Figure 9. A random walk that takes over 6000 steps to return to 0 for the first time.

Since $f_{2n} \asymp 1/n^{3/2}$, the probability that it takes more than N steps for $x(n)$ to first return to zero is comparable to $1/\sqrt{N}$. Thus in 100 trials, one is likely to find a random walk that takes on the order of 10,000 steps to first return to zero. One such walk is shown in Figure 9.

Reflection arguments. The probability of a return to $x = 0$ is closely related to the probability of avoiding the negative integers.
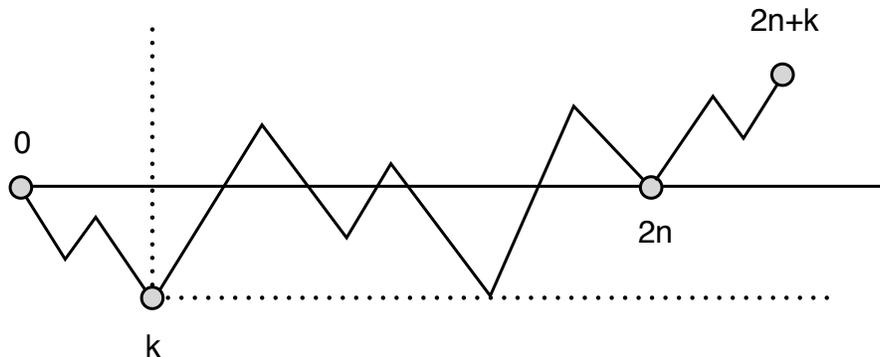
Figure 10. A path that returns to 0 at time $2n$ gives rise to a path that stays on or above the dotted horizontal line.

Theorem 8.7 *The probability that $x(i) \geq 0$ for $i = 0, 1, 2, \ldots, 2n$ is the same as the probability that $x(2n) = 0$.*

The proof is a typical reflection argument. We will see several such arguments in the sequel. It is useful to think of the random walk as a path $(i, x(i))$ in $\mathbb{Z}^2$.

Proof. (Nelson.) We will show that the two types of paths are in bijection. Consider a path from $(0, 0)$ to $(2n, 0)$. Suppose the leftmost minimum of this path occurs at $M = (k, -m)$. Remove the path from $(0, 0)$ to $(k, -m)$, reflect it through a vertical line, and then attach it to the endpoint $(2n, 0)$. (See Figure 10). The result is a path from M to $(2n + k, m)$. Changing the origin to M gives a path with $x(i) \geq 0$ for $i = 1, 2, \ldots, 2n$ (in fact, with $x(2n) = 2m$). The process can be reversed by setting $m = x(2n)/2$ and looking at the last moment when $x(i) = m$. ∎

Equation (8.3) is a consequence of the *ballot theorem*:

Theorem 8.8 *Let $p > q$ be the number of ballots cast for a pair of candidates. Then the probability that the winner is always in the lead as the votes are counted is $(p - q)/(p + q)$.*

Corollary 8.9 *The probability that $x(i) > 0$ for $i = 1, 2, \ldots, n$, given that $x(n) = y > 0$, is y/n.*

Corollary 8.10 *The probability that $x(i) \neq 0$ for $i = 1, 2, \ldots, 2n - 1$, given that $x(2n) = 0$, is $1/(2n - 1)$.*

Proof. Let A be the event $x(i) \neq 0$ for $i = 1, 2, \ldots, 2n - 1$. Then

$$P(A|x(2n - 1) = 1) = 1/(2n - 1)$$

by the preceding result since A becomes the condition $x(i) > 0$. Replacing $x(i)$ with $-x(i)$, we find

$$P(A|x(2n - 1) = -1) = 1/(2n - 1)$$

as well. But if $x(2n) = 0$, then $x(2n - 1) = \pm 1$, so we are done. ∎

Since $P(x(2n) = 0) = u_{2n}$, the previous Corollary implies:

Corollary 8.11 *We have $f_{2n} = u_{2n}/(2n - 1)$.*

Proof of the ballot theorem. We again use a reflection argument. Let $n = p + q$ be the total number of ballots. Think of the ballot count as determining a random walk $(i, x(i))$ in $\mathbb{Z}^2$, with p positive steps and q negative steps, from position $(0, 0)$ to $(p + q, p - q) = (n, p - q)$.

The total number of walks of length n, ending at $x(n) = p - q > 0$, is

$$T = \frac{n!}{p!q!}.$$

We want to know what fraction of these are *positive*, meaning they satisfy $x(i) > 0$ for $i > 0$. Instead we observe that the number of walks that are *not positive* is given by $2B$, where

$$B = \frac{(n - 1)!}{p!(q - 1)!}.$$

To see this, first observe that B is the number of walks with $x(1) = -1$ and $x(n) = p - q$. (Once this first negative step has been taken, the walk continues with p positive steps and $q - 1$ negative steps.) These walks are certainly not positive. The other way a walk can fail to be positive is if it starts out $x(1) = 1$, but then $x(i) = 0$ for some $i > 1$. Consider the least such i. Replacing $x(j)$ with $-x(j)$ from time 0 up to time i, we obtain a

walk from $(1, -1)$ to $(n, p - q)$. So the number of non-positive walks with $x(1) = 1$ is *also* B.

Since
$$\frac{B}{T} = \frac{q}{n} = \frac{q}{p + q},$$
we have
$$\frac{T - 2B}{T} = \frac{p - q}{p + q},$$
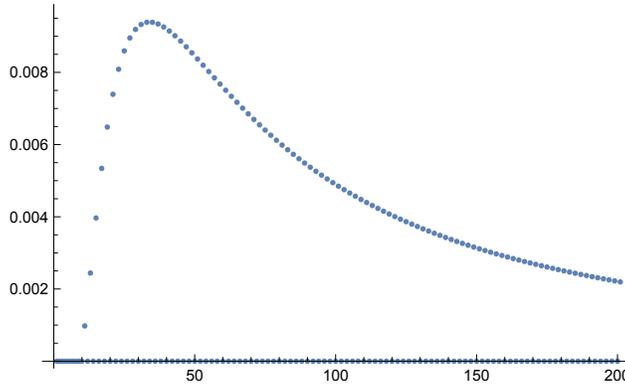and this is the probability that a path ending at $(n, p - q)$ is positive. ■

Figure 11. Probability that $x(n) = 10$ and $x(i) < 10$ for $i < n$.

First passage: distribution in time. The central limit theorem records the distribution of $x(n)$ at a fixed time n. One can also look at the distribution in *time* of the event that $x(n)$ makes its *first visit* to a particular site $x = m$. This random variable is defined by

$$\tau(m) = \inf\{n \geq 0 \ : \ x(n) = m\}.$$

Since $x(n)$ is recurrent, $\tau(m)$ is finite almost surely.

For an example, see Figure 11. We note that $P(\tau(m) = n)$ is exponentially small until n is comparable to m^2; it then has a $1/n^{3/2}$ tail.

Here are some useful principles for studying $\tau(m)$.

1. It is easy to see that

$$P(\tau(1) = 2n - 1) = f_{2n},$$

   the probability that the first return to 0 occurs at time $2n$.

2. Let $\tau_1(1), \ldots, \tau_m(1)$ be independent random variables with the same distribution as $\tau(1)$. Then $\tau(m)$ has the same distribution as $\sum_1^m \tau_i(1)$.

   To see this observe that to reach m, $x(n)$ must first reach 1 for the first time, then reach 2, etc.

3. By definition, the generating function for $\tau(1)$ is given by

$$F(z) = E(z^{\tau(1)}) = \sum_0^\infty P(\tau(m) = n)z^n.$$

   The preceding observation shows that the generating function for $\tau(m)$ is given by $F(z)^m$.

4. In fact, the generating function $F(z)$ for $\tau(1)$ satisfies

$$F(z) = (1/2)(z + zF(z)^2),$$

   and thus it is given by:

$$F(z) = \frac{1 - \sqrt{1 - z^2}}{z}.$$

   To prove the first formula for $F(z)$, observe that either $x(1) = 1$ or $x(1) = -1$, each with probability one half. In the first case $\tau(1) = 1$; the generating function for this distribution is z. In the second case, $x(n)$ must now climb from -1 to $+1$; the total time required has the same distribution as $1 + \tau(2)$, which has generating function $zF(z)^2$.

5. The next result relates the distribution of $\tau(m)$ in time to the distribution of $x(n)$ in space; it thus allows us to use the central limit theorem and allied results to estimate $\tau(m)$. The proof is another reflection argument.

   Theorem 8.12 *For any $n \geq 0$, we have:*

$$P(\tau(m) \leq n) = P(x(n) = m) + 2P(x(n) > m). \qquad (8.4)$$

   Proof. To see this, consider the path $(i, x(i))$ up to time $i = n$, and assume $x(i) = m$ for some $i \leq n$. Then we can reflect that path

95

from time i to time n through the line $x = m$ to obtain a new and different path $x'(n)$, also with $x(i) = m$. Now if $x(n) = m$, we also have $x'(n) = m$, so we have just permuted the set of paths with $x(n) = m$. On the other hand, if $x(n) < m$, then $x'(n) > m$. So we have shown the paths with $x'(n) < m$ can each be matched to one with $x'(n) > m$. And of course all the paths with $x(n) \geq m$ satisfy $\tau(m) \leq n$. So the sum of the two types gives the desired probability. ∎

Corollary 8.13 *For $k \geq 1$, we have $P(\tau(m) \leq m^2/k^2) = O(2^{-k})$.*

Proof. Let $n = m^2/k^2$, so $m = k\sqrt{n}$. By equation (8.4), the probability that $\tau(m) \leq n$ is bounded by

$$2P(x(n) \geq m) = 2P(x(n) \geq k\sqrt{n}) = O(2^{-k})$$

by Corollary 8.3 . ∎

9   The continuum limit I: Harmonic functions

It is intuitively clear that as $h \to 0$, the finer and finer lattices $h\mathbb{Z}^d$ converge to the *continuum* $\mathbb{R}^d$. How do random walks and discrete harmonic functions behave in the limit?

In this section we will show that discrete harmonic functions converge to smooth ones. In particular, the Dirichlet problem on Ω can be solved by first approximating Ω by a grid, and then passing to the limit. This method is practical for numerical analysis.

The theory has a long history, going back to the work [CFL] of Courant, Friedrichs and Lewy in 1928; see also [Fer]. Our treatment is based in part on [DS, Ch.4].

Discrete calculus.   For convenience we will just treat the case $d = 2$; the arguments generalize to all dimensions d. To emphasize the eventual connection to complex analysis, we will write a random walk on $\mathbb{Z}^2 = \mathbb{Z}[i]$ as

$$z(n) = x(n) + iy(n).$$

Let $\Omega \subset \mathbb{C}$ be an open set, let $h > 0$ be a mesh size, and let

$$\Omega_h = \Omega \cap h\mathbb{Z}[i]$$

be the corresponding discrete approximation to Ω.

Given a function

$$f : \Omega_h \to \mathbb{R},$$

we define its discrete derivative in the x–direction by:

$$D_x f(z) = \frac{f(z+h) - f(z)}{h}, \tag{9.1}$$

and similarly for D_y. We also define

$$D_x^- f(z) = \frac{f(z) - f(z-h)}{h},$$

and then define

$$Af(z) = (1/4)(D_x D_x^- + D_y D_y^-)(f).$$

We then have

$$Af(z) = \frac{1}{h^2}\left(\frac{(f(z+h) + f(z-h) + f(z+ih) + f(z-ih))}{4} - f(z)\right).$$

In other words, $Af(z)$ is the usual discrete Laplacian $P - I$, up to a factor of h^2. As usual, we say f is harmonic if $Af = 0$ on the points of Ω with 4 neighbors.

Compactness. Now let $h_n \to 0$ be a sequence of mesh sizes converging to zero, and for brevity let

$$\Omega_n = \Omega \cap h_n\mathbb{Z}[i]. \tag{9.2}$$

We will show:

Theorem 9.1 *Let $f_n : \Omega_n \to [-1, 1]$ be a sequence of bounded, discrete harmonic functions. Then after passing to a subsequence, there exist a harmonic function*

$$F : \Omega \to [-1, 1]$$

such that

$$\sup_{z \in K \cap h_n\mathbb{Z}[i]} |f_n(z) - F(z)| \to 0$$

for every compact set $K \subset \Omega$.

In brief, we say $f_n \to F$ uniformly on compact sets.

Using the theorem above, we will also sketch the proof the Dirichlet problem can be solved by first discretizing and then passing to the continuum limit.

PL extension. The statement $f_n \to F$ can also be formulated as follows. Think of the lattice $h_n \mathbb{Z}^2$ as the vertices of a tiling of $\mathbb{R}^2$ by squares of side length h_n. Subdivide each square into two triangles along the diagonal. Now extend f_n to $\mathbb{R}^2$ so it is linear on each triangle, and zero on each vertex $z \in h_n \mathbb{Z}^2 - \Omega$.

Then the conclusion of the theorem is that these extended PL functions satisfy

$$f_n | \Omega \to F$$

uniformly on compact subsets of Ω.

Hitting a line. The next result is the key technical point.

Theorem 9.2 *Let $z(n) = x(n) + iy(n)$ be a random walk on $\mathbb{Z}[i]$ starting at $z(0) = i$. Then the probability that $z(n)$ exits the square $[-N, N] \times [-N, N]$ before it hits the real axis is $O(1/N)$.*

Proof. Let τ denote the least n such that $z(n)$ lies in the boundary of the square. Then $|y(\tau)| = N$ or $|x(\tau)| = N$.

In the first case, we can regard $y(n)$ as a random walk with pauses on $[0, N]$, start at $y(0) = 1$. As we have seen, the probability that $y(n)$ hits N before it hits 0 is a harmonic function of its starting point, namely $p(y) = y/N$. Thus the probability that $|y(\tau)| = N$ is bounded by $1/N$.

In the second case, we can regard $x(i)$ as a random walk with pauses on $[-N, N]$ that hits an endpoint at time τ. It is rather likely that $\tau \geq N^2$. In fact, by Corollary 8.13, $P(\tau \leq N^2/k^2) = O(2^{-k})$. (The pauses only increase the value of τ.)

We now examine what happens with $y(n)$ during this walk. Let S denote the number of true steps in the random walk $y(n)$ up to time τ (pauses are ignored). The probability that $S < \tau/3$ is very low (much less than $1/N$), so we may assume S is comparable to τ. Now the probability that a random walk starting at 1 does not visit 0, up to time S, is $O(1/\sqrt{S})$ (cf. Theorem 8.7). Thus the probability of not hitting the real axis is $O(1/\sqrt{\tau})$, and this is $O(1/N)$ if $\tau \geq N^2$.

Now it is also possible that τ is comparable to N^2/k^2, for some integer $k \geq 2$. The probability of this event is $O(2^{-k})$, and the probability that $y(i)$ does not hit zero up to time N^2/k^2 is $O(k/N)$. Thus the contribution of these remaining cases to the probability of not hitting the real axis is bounded by a multiple of

$$\sum_k 2^{-k}(k/N) = O(1/N).$$

This completes the proof. ∎

With more work, the same idea can be used to control the probability that $z(n)$ crosses a line with any given slope. See also the comment on the function $\arg(z)$ below.

Combinatorial derivatives. The key to the proof of Theorem 9.1 is the following *a priori* bound on the gradient of a discrete harmonic function.

Theorem 9.3 *Let $f : \mathbb{Z}[i] \to [-1,1]$ be harmonic on the square $[-N, N] \times [-N, N]$. Then*

$$|f(i) - f(-i)| \leq C/N$$

for a universal constant $C > 0$.

Proof. (Following [DS, Ch.4].) Let $z_+(n)$ denote a random walk started at $z_+(0) = +i$ and reaching the boundary of the square at time τ_+. Similarly define $z_-(n)$ and τ_-. Then

$$f(\pm i) = E(f(z_\pm(\tau_\pm))). \tag{9.3}$$

We would like to match each path $z_+(n)$ to a corresponding path $z_-(n)$. This is done as follows. Let σ be the first moment when $z_+(n)$ hits the real axis. Note that σ is finite with probability one, since $z_+(n)$ is recurrent. By reflecting the path up to time σ through the real axis, we obtain a path of the form $z_-(n)$. This map between paths is a measure–preserving bijection.

By construction, $z_-(n) = z_+(n)$ for $n \geq \sigma$. Thus $\tau_+ = \tau_-$ provided $\tau_+ < \sigma$. In this case, we also have

$$f(z_+(\tau)) = f(z_-(\tau)),$$

so both paths make the same contribution to the value of $f(\pm i)$ in equation (9.3).

99

This almost proves that $f(i) = f(-i)$. The only issue is that sometimes $\sigma > \tau_+$. But this event has probability $O(1/N)$, by Theorem 9.2. Since f is bounded, the total contribution from these path is $O(1/N)$. Thus $|f(i) - f(-i)| = O(1/N)$. ∎

Corollary 9.4 *Under the same hypothesis, we have*

$$|f(0) - f(i)| = O(1/N).$$

Proof. The preceding argument shows that within $B(0, 10)$, we can regard $f(z)$ as constant on the 4 cosets of $2\mathbb{Z}[i]$, up to an error of $1/N$. Since $f(z)$ is harmonic, its value on one coset is the average of its values over 2 other cosets. Thus all these values are the same, to within an error of $O(1/N)$. ∎

Note: the preceding argument is related to the fact that a harmonic function on $\mathbb{Z}[i]$, invariant under translation by $2\mathbb{Z}[i]$, is constant.

Corollary 9.5 *Under the same hypothesis, we have*

$$|f(z) - f(w)| \leq C|z - w|/N$$

provided $|z|, |w| \leq N/2$.

Proof. Connect z to w by a path z_i of length $O(|z - w|)$ with $|z_i - z_{i+1}| = 1$ and apply the preceding bound. ∎

Corollary 9.6 *If $f : \Omega_h \to [-1, 1]$ is harmonic, then we have*

$$|D_x f|, |D_y F| \leq C(K)$$

on each compact set $K \subset \Omega$.

Proof. On K there exists an $r > 0$ such that $B(z, 10r) \subset \Omega$ for all $z \in K$. Thus each $z \in K \cap \Omega_h$ is contained in a combinatorial square of side length $2N$ comparable to $1/h$. Thus the values of f at adjacent sites in Ω_h differ by $O(h)$, so $|D_x f|$ and $|D_y f|$ are $O(1)$. ∎

Proof of Theorem 9.1. By the preceding Corollary, the PL extensions of the functions $f_n(z)$ are uniformly Lipschitz on any compact set $K \subset \Omega$. Thus after passing to a subsequence, there is a continuous function $F : \Omega \to \mathbb{R}$ such that $f_n \to F$ uniformly.

Now consider the functions $D_x f_n$ defined by equation (9.1). By Corollary 9.5, $D_x f_n$ is uniformly bounded by $C(K)$ on any given compact set $K \subset \Omega$. Even better, $D_x f_n$ is a discrete harmonic function. Thus after passing to a subsequence, we can assume $D_x f_n \to G$ for some *continuous* function G. Now at a small enough scale, $G(z)$ is nearly constant. The implies that $f_n(z)$ is nearly linear at the same scale. Passing to the limit, it is straightforward to show that dF/dx exists and $G(z) = dF/dx$. The same argument applies with respect to y.

Since the particular derivatives dF/dx and dF/dy are both continuous, F is differentiable. It follows similarly that F has derivatives of all orders. We also have

$$0 = Af_n \to (1/4)\Delta F$$

uniformly on compact sets, so F is harmonic. ∎

Remark on differentiability. Because $h_n \mathbb{Z}^2$ is invariant under horizontal and vertical translations, it is easiest (in the proof above) to control dF/dx and dF/dy individually, show they are continuous, and then conclude that F is differentiable.

We note that the hypothesis of continuous derivatives is essential. For example, the continuous function $f(z) = xy/|z|$ is not differentiable at the origin, even though both its derivatives exist everywhere. Away from zero, its derivatives are given by

$$df/dx = y^3/|z|^3 \quad \text{and} \quad df/dy = x^3/|z|^3,$$

and these homogeneous functions are not continuous at the origin.

Superharmonic functions and $\arg(z)$. The standard branch of $\arg(z)$ lies in $(-\pi, \pi]$. This function is harmonic in the open first quadrant Q of the hyperbolic plane. Using this fact, one can show that the harmonic measure of the imaginary axis I satisfies

$$\omega(z, I, Q) = \arg(z)/\pi.$$

If $\arg(z)$ were also harmonic in $\mathbb{Z}^2$, it could then be used to give a good estimate for the hitting probabilities considered above. It fact it is nearly

harmonic; we have
$$A \arg(z) = O(|z|^{-4})$$
when $|z|$ is large. And indeed, $\arg(z)$ is superharmonic when it lies in the range $[0, \pi/4]$. This fact can be exploited to give another proof of Theorem 9.2.

Solving the Dirichlet problem. Finally we show the Dirichlet problem can be solved by discrete approximation. This method is quite practical for numerical analysis (although finite elements and sparse linear algebra, rather than random walks on a lattice, are used in the algorithms).

Effect of discretization on the Laplacian. To begin the discussion, let $f\Omega \to \mathbb{R}$ be a smooth function, and let f_h be the restriction of f to

$$\Omega_h = \Omega \cap h\mathbb{Z}[i].$$

Now for a quadratic polynomial $q(z)$, we have

$$Aq(z) = (1/4)\Delta q.$$

By Taylor's theorem, we can approximate $f(z)$ by such a polynomial with error $O(h^3)$ between adjacent lattice points, and thus

$$Af(z) = (1/4)\Delta f(z) + O(h), \tag{9.4}$$

uniformly on compact subsets of Ω.

Point of departure. For simplicity we treat the case where we already have a smooth solution to the Dirichlet problem. In detail, suppose:

1. $\Omega \subset \mathbb{C}$ is a compact domain with smooth boundary;

2. $f : \overline{\Omega} \to \mathbb{R}$ is harmonic on Ω; and

3. f extends to a smooth function on a neighborhood of Ω.

With more work we could allow Ω to be a Jordan domain, and simply require that f is continuous on $\overline{\Omega}$, but the smooth case goes to the heart of the matter, gives good error bounds and illustrates a useful technique.

Let $h_n \to 0$ be a sequence of mesh size, define Ω_n by (9.2), and define

$$f_n : \partial\Omega_n \to \mathbb{R}$$

so that

$$|f_n(x) - f(y)| = O(h_n) \tag{9.5}$$

whenever $x \in \partial\Omega_n$, $y \in \partial\Omega$ and $|x - y| \leq h_n$.

Let $f_n : \overline{\Omega}_n \to \mathbb{R}$ be the unique discrete harmonic function with the given values on $\partial\Omega_n$. We will show:

Theorem 9.7 *As $n \to \infty$, $f_n \to f$ uniformly. In fact*

$$\sup_{\overline{\Omega}_n} |f_n(x) - f(x)| = O(h_n).$$

We remark that an error of size at least h_n is almost inevitable, since one must approximate points of $\partial\Omega$ by points in the lattice $h_n\mathbb{Z}^n$.

Proof of Theorem 9.7. (Schramm.) The main idea in the proof is that the known smooth solution $f(x)$ allows us to produce a nearby discrete function g_n that controls f_n. First, we extend f smoothly to a neighborhood of $\overline{\Omega}$, so we can evaluate it on $\overline{\Omega}_n$ once n is large enough. Next, we set

$$g_n(z) = f(z) + a_n|z|^2 - b_n$$

on $\overline{\Omega}_n$, where $a_n, b_n \geq 0$ are chosen so that

$$A(g_n) \geq 0 \quad \text{and} \quad g_n(x) \leq f_n(x) \quad \forall x \in \partial\Omega_n. \tag{9.6}$$

The first condition says that g_n is subharmonic. In view of equation (9.4), and the fact that $\Delta|z|^2 = 4$, we can achieve $A(g_n) \geq 0$ with $a_n = O(h_n)$. Then, in view of equation (9.5), the second condition can be achieved with $b_n = O(h_n)$ as well. Thus

$$\max_{\overline{\Omega}_n} |g_n - f| = O(h_n).$$

On the other hand, since f_n is a discrete harmonic functions, the two conditions (9.6) guarantee that:

$$g_n(x) \leq f_n(x) \quad \forall x \in \overline{\Omega}_n.$$

Using the same idea, we can construct a superharmonic function $G_n = f + O(h_n)$ such that $f_n \leq G_n$. Thus $g_n \leq f_n \leq G_n$ and both g_n and G_n are small modifications of f; consequently

$$\max_{\overline{\Omega}_n} |f_n - f| = O(h_n)$$

as well. ∎

103

10 The continuum limit II: Brownian motion

In addition to studying the behavior of discrete harmonic functions in the limit $h_n \mathbb{Z}^d \to \mathbb{R}^d$, we can also study the limiting behavior of random walks.

Suitably rescaled in both space and time, random walks $x(n)$ on the $\mathbb{Z}^d$ lattice converge to a continuous process $B(t)$ on $\mathbb{R}^d$, called *Brownian motion*. In this section we will describe several approaches to describing this limit and its symmetry, especially for $d = 1$ and $d = 2$. We will then explain how $B(t)$ can be used to solve the Dirichlet on a domain $\Omega \subset \mathbb{C}$ directly, and how it provides a new geometric and dynamical perspective on several conformal invariants.
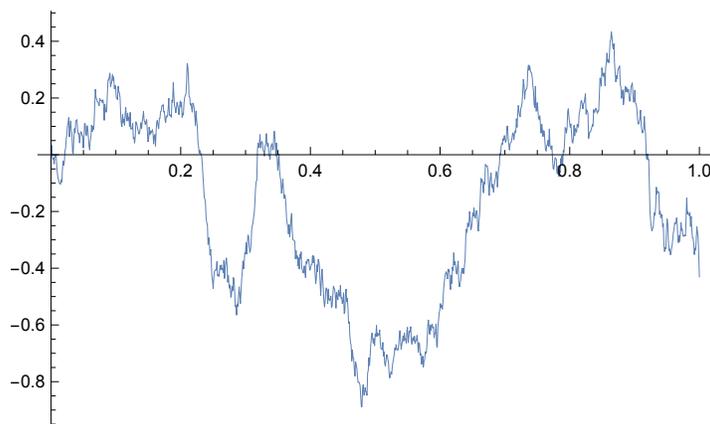

Figure 12. An instance of Brownian motion $B : [0, 1] \to \mathbb{R}$.

Axioms for Brownian motion. We start with Brownian motion $B(t)$ on $\mathbb{R}$. This motion is described by a probability measure μ on $C(\mathbb{R})$, with the following characteristic properties:

1. We have $B(0) = 0$ for almost every $B \in C(\mathbb{R})$.

2. For any $t_0 < t_1 < \cdots < t_n$, the increments $B(t_{i+1}) - B(t_i)$ are independent.

3. The increment $B(t) - B(s)$ has mean zero and variance $|t - s|$.

4. Each increment is a Gaussian random variable.

By the third axiom above, we have $E(B(t)^2) = t$. This implies that $|B(t)|$ is usually of size $\sqrt{t}$, and thus there is no chance that $B(t)$ is differentiable or even Lipschitz at the origin. The same applies at every other point.

The axioms imply that $B(t)$ has the Gaussian distribution

$$G_t(x) = \frac{1}{\sqrt{2\pi t}} \exp(-x^2/2t),$$

with standard deviation $\sqrt{t}$. The same is true for $B(s+t) - B(s)$. In fact, since

$$B(t) = \lim_{n\to\infty} \sum_1^n X_i,$$

where $X_i = B(it/n) - B((i-1)t/n)$ are independent random variables with mean zero and variance $1/n$, the statement that $B(t)$ is Gaussian almost follows from the central limit theorem. (E.g. it would be enough to add some control over $E(X_i^4)$.)

The main result is:

Theorem 10.1 *There exists a unique measure on $C(\mathbb{R})$ satisfying the axioms above.*

The resulting random process is called *standard* Brownian motion on $\mathbb{R}$. The corresponding measure on $C(\mathbb{R})$ is *Wiener measure*.

Natural approaches to the construction of $B(t)$ include the following:

1. Taking the scaling limit of discrete random walks;

2. Take a limit of measures on piecewise–linear functions over finer and finer grids; and

3. Constructing $B'(t)$ first as white noise, and then integrating.

In all cases a pivotal step is to show that $B(t)$ is (almost surely) continuous. In fact $B(t)$ is α–Hölder continuous for every exponent $\alpha < 1/2$; it almost has $1/2$ a derivative in L^2.

Here is a version of the first construction. Consider a random walk $x(i)$ on $\mathbb{Z}$. Extended each walk to a piecewise–linear function $x(s)$ on $\mathbb{R}$. Then define a probability measure μ_n on $C(\mathbb{R})$ supported on functions of the form

$$B_n(t) = \frac{x(nt)}{\sqrt{n}},$$

by taking $x(i)$ to be random walk as usual. Note that $B_n(1) = x(n)/\sqrt{n}$ satisfies $E(B_n(1)^2) = 1$. We then have

Theorem 10.2 (Donsker) *The measures μ_n converge to Wiener measure as $n \to \infty$.*

In the sequel we will discuss the other two constructions of Brownian motion; the same considerations underly the proof of Donsker's theorem.

Remarks on Brownian motion. Before turning to the construction of Brownian motion, we make a several remarks.

1. The axioms for Brownian motion immediately describe, for any $t_1 < t_1 < \cdots < t_n$, the distribution of $(B(t_1), B(t_2), \ldots, B(t_n))$ in $\mathbb{R}^n$. Since $B(t)$ is continuous, there is at most one measure on $C(\mathbb{R})$ with the prescribed joint distributions.

2. Although $B(t)$ is very likely to satisfy $|B(t)| \leq 10\sqrt{t}$, for $t > 0$ there is a positive probability that $B(t) \in (a, b)$ for *any* interval (a, b). Unlike a random walk, Brownian motion can move arbitrarily fast, albeit with very small probability.

   This behavior is related to the fact that the Gaussian distribution is real–analytic, so it cannot vanish on an interval.

3. Brownian motion provides a solution to the heat equation

$$\frac{df_t}{dt} = (1/2)\Delta f$$

   on $\mathbb{R}$, with given initial data f_0, namely:

$$f_t(x) = E_x(f(B(t))) = (G_t * f_0)(x)$$

   for $t > 0$. Here the expectation is taken with respect to a Brownian path with $B(0) = x$. The relation is immediate from the fact that $G_t(x)$ is a fundamental solution to the heat equation.

4. For any $\lambda > 0$, the function $B(\lambda^2 t)/\lambda$ is also a copy of standard Brownian motion. Informally,

$$\lambda^{-1}B(\lambda^2 t) = B(t),$$

in the sense that both sides define the same measure on $C(\mathbb{R})$.

To see this, the main point is to check the variance; the variance of $B(\lambda^2 t)$ is $\lambda^2 t$, and this is corrected by the scale factor of λ^{-1}. The statement then follows from uniqueness of Brownian motion.

5. One can define Brownian motion on $\mathbb{R}^d$ by $B(t) = (B_1(t), \ldots, B_d(t))$, where $B_i(t)$ are *independent* Brownian motions on $\mathbb{R}$. Then the variance of $B(t) - B(s)$ is $d|t - s|$; in particular we have

$$E(B(t)^2) = d|t|.$$

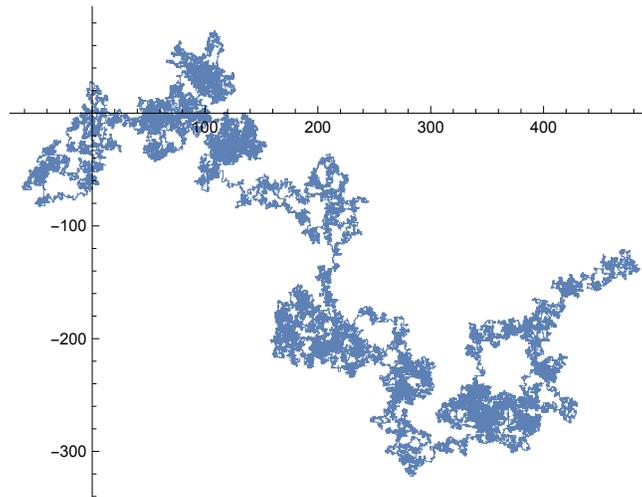See Figure 13 for an example in $\mathbb{R}^2$.

Figure 13. Brownian motion in the plane.

6. Although we have used a basis for $\mathbb{R}^d$ to define $B(t)$, the process is in fact invariant under rotation. To see this, we note that the distribution of $B(t)$ on $\mathbb{R}^d$ is given by the spherically symmetric Gaussian

$$G_t^d(x) = (2\pi t)^{-d/2} \exp(-|x|^2/t) = \prod_1^d (2\pi t)^{-1/2} \exp(-x_i^2/t).$$

The key point here is that a *product* of Gaussians distributions is again *Gaussian*.

7. For a random variable X on $\mathbb{R}^d$, it is more natural to study its *co-variance matrix* $C(X) = E(X_i X_j)$, rather than the variance $E(|X|^2)$. The covariance matrix for $X = B(t) \in \mathbb{R}^d$ is given by tI. Its *trace* is $E(|X|^2)$.

8. Because a random walk $x(i)$ on $\mathbb{Z}^d$ changes only one coordinate at a time, to obtain Brownian motion under rescaling once must accelerate the time coordinate by a factor of d. Then the walks $B_n(t) = x(dnt)/\sqrt{n}$ converge to Brownian $B(t)$ on $\mathbb{R}^d$.

9. It is remarkable that a random walk on the lattice $\mathbb{Z}^d$ converges to a rotationally–invariant process $B(t)$ on $\mathbb{R}^d$. The coordinate directions of the underling lattice disappear in the limit. This is a kind of universality.

   It is also true that $B(t)$ is invariant under similarities, up to a time change; that is, the *paths* $\lambda B(t)$ have the same distribution as the paths $B(t)$, for any $\lambda > 0$. Thus we have a kind of conformal invariance of the continuum limit of random walks.

10. We note that Brownian motion *per se* is *not* conformally invariant; we have used the Euclidean metric to set the pace of motion.

    In dimension $d = 2$, we can also define Brownian motion B_ρ on $(\mathbb{C}, \rho)$ for any conformal metric ρ. To do so, we define a new time parameter by

    $$s = \int_0^t \rho(B(u))^2 \, du,$$

    and then set $B_\rho(s) = B(t)$.

11. In the complex plane, an even stronger conformal invariance holds: if $f : U \to V$ is a conformal map between two regions in $\mathbb{C}$, and $w = f(z)$, then f sends standard Brownian paths starting at z to standard Brownian paths starting at w, up to a time change. In fact, the pullback of Brownian motion on V to U is just Brownian motion on U with respect to the metric $\rho(z) = |f'(z)| \, |dz| = f^*(|dz|)$.

12. Even more remarkably, if $f : (U, 0) \to (\mathbb{C}, p)$ is a holomorphic function, perhaps with branch points and perhaps not injective, it is still true that $f(B(t))$ is a time–changed copy of Brownian motion on $\mathbb{C}$, started

at p. This motion will usually not be defined for all time; it stops when $B(t)$ reaches ∂U.

As one example, the map $f : \mathbb{C} \to \mathbb{C}^*$ given by $f(t) = \exp(t)$ sends Brownian motion starting at 0 to Brownian motion starting at $f(0) = 1$. In this case the path $f(B(t))$ is defined for all time. It also *avoids $z = 0$*. This proves:

> *The probability that a Brownian path $B(t)$ in $\mathbb{C}$ passes through a given point $p \neq B(0)$ is zero.*

Construction of 1–dimensional Brownian motion I: Lévy's method. We now turn to the construction of Brownian motion $B : [0,1] \to \mathbb{R}$ with $B(0) = 0$. One can readily obtain Wiener measure on $C(\mathbb{R})$ once it is known on $C[0,1]$.

We will specify $B(t)$ for t in the ring of *dyadic rationals* $\mathbb{Z}[1/2] = \bigcup 2^{-n}\mathbb{Z}$. Every dyadic rational can be expressed in the form

$$t = \frac{p}{2^n}$$

with $p \in \mathbb{Z}$. The dyadic rationals with denominator 2^n cut $[0,1]$ into 2^n *standard dyadic intervals*, each of the form $[p, p+1]/2^n$.

Let ξ_t denote a family of independent Gaussian random variables, indexed by the dyadic rationals in $[0,1]$. Using these variables, we define a sequence of continuous approximations $B_n(t)$ to $B(t)$. The approximations are characterized by the following properties:

1. We have $B_0(0) = 0$ and $B_0(1) = \xi_1$.

2. The function $B_n(t)$ is linear on each dyadic interval of length 2^{-n}. Thus it is uniquely determined by its values $B_n(p/2^n)$, $p = 0, 1, \ldots, 2^n$.

3. We have $B_n(t) = B_{n-1}(t)$ if $t = p/2^n$ with p even.

4. We have
$$B_n(t) = B_{n-1}(t) + \frac{1}{\sqrt{2^{n+1}}}\, \xi_t,$$

   if $t = p/2^n$ with n odd.

See Figure 14 for an instance of this construction, which was used to produce Figure 12.

The benefit of this construction is that the values of $B(t)$ are consistently defined a sequence of denser and denser points. The desired independence of increments between nth level dyadic points is readily verified for $B_n(t)$. We will check the variance of increments, and then turn to the main issue, which is uniform continuity.
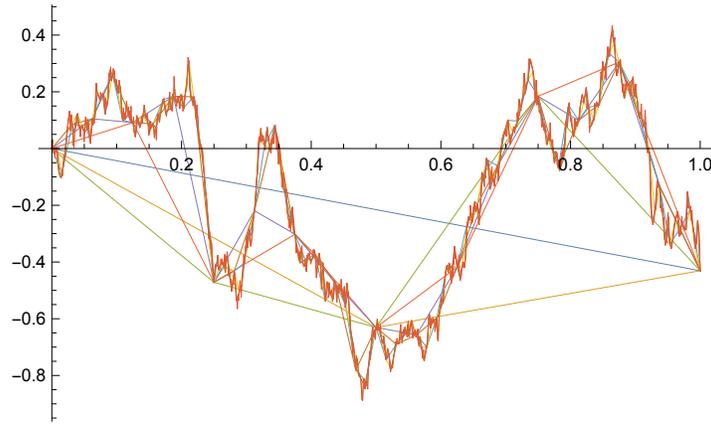

Figure 14. Construction of Brownian motion.

Variance. We explain why, by induction, if $[a, b] = [p, p+1]/2^n$ is a standard dyadic interval in $[0, 1]$ at level n, the increment

$$D_n(a, b) = B_n(a) - B_n(b)$$

has variance $|a - b| = 2^{-n}$. The case $n = 0$ is immediate. Let $t = (a + b)/2$ be the midpoint of an interval (a, b) at level $n - 1$. Then

$$D_n(a, t) = \frac{1}{2} D_{n-1}(a, b) + \frac{1}{\sqrt{2^{n+1}}} \xi_t.$$

The first term has, by induction, variance $(1/4)2^{-(n-1)} = 2^{-(n+1)}$. The variance of the second term is the same. By independence, the variance of $D_n(a, t)$ is their sum, 2^{-n}, as desired. The same reasoning applies to $D_n(x, b)$.

Continuity. We now define, for every dyadic rational $t \in [0, 1]$, $B(t) = \lim B_n(t)$. This sequence is eventually constant.

We claim that $B(t)$ extends to a continuous function on $[0, 1]$. More precisely, fixing $0 < \epsilon$, for almost every $B(t)$ there exists a C such that

$$|B(t) - B(s)| \leq C|t - s|^{1/2-\epsilon}. \tag{10.1}$$

for any dyadic $s, t \in [0, 1]$. Consequently $B(t)$ has a continuous extension and moreover:

Theorem 10.3 *Brownian motion $B : [0, 1] \to \mathbb{R}$ is almost surely Hölder continuous of exponent $1/2 - \epsilon$, for all $\epsilon > 0$.*

Lemma 10.4 *Suppose equation (10.1) holds for all standard dyadic intervals $[s, t]$ of length 2^{-n}. It then holds, with a different constant C, for all pairs of dyadic numbers $s, t \in [0, 1]$.*

Proof of Theorem 10.3. Let $[s, t]$ be a basic dyadic interval of length 2^{-n}. Let p_n be the probability that (10.1) is violated with $C = 1$. Recall that the standard deviation of $B(t) - B(s)$ is $\sigma = \sqrt{|s - t|}$. Thus when this equation is violated, the number of standard deviations required is $|t - s|^{-\epsilon} = 2^{n\epsilon}$. Since this variable is normally distributed, we have

$$p_n = O(\exp(-2^{n\epsilon})).$$

Here we using the soft bound for mean zero Gaussian random variables,

$$P(|X| > k\sigma) = O(2^{-k})$$

and in fact 2 can be replaced by any $\Lambda > 0$.

Now there are 2^n basic dyadic intervals of length 2^{-n}. Thus expected number of these intervals where (10.1) fails, with $C = 1$, is bounded by a multiple of

$$\sum_n 2^n \exp(-2^{n\epsilon}) = \sum \exp(an - b^n)$$

for suitable $a > 0$ and $b > 1$. The $-b^n$ term dominates, and therefore this sum is finite.

Consequently (10.1) holds with $C = 1$ for all but finitely many standard dyadic intervals $[s, t]$. Thus we can choose C so it holds for all standard dyadic intervals. ∎

Construction of Brownian motion II: White noise. We now turn to a different construction of Brownian motion, using a version of the Gaussian distribution on Hilbert space.

For the convenience of using Fourier series, we will carry out the construction of Brownian functions and white noise on the unit length circle $S^1 = \mathbb{R}/\mathbb{Z}$. Then $L^2(S^1)$ has a complex orthonormal basis given by z^n, $n \in \mathbb{Z}$, with $z = \exp(2\pi i x)$, $x \in \mathbb{R}$. This basis gives an isomorphism $L^2(S^1) \cong \ell^2(\mathbb{Z})$, given by

$$f(z) = \sum a_n z^n$$

with $\|f\|^2 = \sum |a_n|^2$.

We can then define, for $s \in \mathbb{R}$, the Sobolev spaces H_s as formal sums as above with

$$\sum (1 + |n|^2)^s |a_n|^2 = \|f\|_{H_s}^2 < \infty$$

or more concretely as sequences (a_n) satisfying this summability condition. When $s > 0$ is an integer, H_s consists of functions with s (distributional) derivatives in L^2; and H_{-s} is a Hilbert space of distributions smooth enough to pair with elements of H_s. Note that when $f \in H_s$ we have

$$|a_n| = O(n^{-s}). \tag{10.2}$$

Here are some important examples.

1. $H_0 = L^2(S^1)$.

2. $H_{1/2+\epsilon} \subset C(S^1)$. Indeed, if $f = \sum a_n z^n$ is in H_s, then we have by Cauchy–Schwarz (omitting $n = 0$):

$$\sum |a_n| = \sum |n|^s |n|^{-s} |a_n| \leq \left(\sum |n|^{2s} |a_n|^2 \right)^{1/2} \left(\sum |n|^{-2s} \right)^{1/2},$$

and the final sum is finite when $2s > 1$. (Put differently, the condition on (a_n) puts it into $\ell^1(\mathbb{Z})$).

3. We have $H_1 \subset C^\alpha(S^1)$ with $\alpha = 1/2$, i.e. we have the Hölder continuity condition

$$|f(z) - f(w)| = O(|z - w|^{1/2}).$$

For in this case, $f' \in L^2(S^1)$, and

$$|f(z) - f(w)| \leq \int_z^w 1 \cdot |f'| \leq \|f'\|_2 \cdot |z - w|^{1/2},$$

again by Cauchy–Schwarz.

112

4. More generally, if $f \in H^s$, $s > 1/2$, then f is α–Hölder continuous for every exponent $0 < \alpha < s - 1/2$.

To see this, we note that for any $\alpha \geq 0$ we have

$$\|z^n\|_{C^\alpha} = O(|n|^\alpha)$$

(with $n = 0$ omitted as before), and thus

$$\|f\|_{C^\alpha} = O(\sum |a_n| \cdot |n|^\alpha).$$

By Cauchy–Schwarz we have

$$\sum |n|^s |a_n| \cdot |n|^{\alpha - s} \leq \|f\|_{H_s} \left(\sum |n|^{2\alpha - 2s}\right)^{1/2},$$

and the final sum is finite if and only if $2(\alpha - s) < -1$, which is equivalent to $\alpha < s = 1/2$.

White noise. Brownian motion $B : [0, 1] \to \mathbb{R}$, normalized so $B(0) = 0$, furnishes us with a *random distribution*, i.e. element of the dual of $C^\infty[0, 1]$, by taking $W = B'(t)$. This distribution is characterized by

$$\langle \phi, W \rangle = \int_{[0,1]} \phi W = -\int_0^1 \phi'(t) B(t) \, dt$$

for any smooth test function ϕ.

To construct it formally, let $H = L^2[0, 1]$. Then H is abstractly a separable Hilbert space. Choose an orthonormal basis $e_n \in H$ and a sequence of independent standard Gaussian random variables ξ_n. Then formally, a sample of the 'Gaussian distribution' on H give a random vector

$$W = \sum \xi_n e_n.$$

Now we have

$$E(\|W\|^2) = \sum E(|\xi_n|^2) = \infty,$$

so W does not make sense as an element of H. However, for any given $f = \sum a_n e_n \in H$, we have $\|f\|^2 = \sum |a_n|^2 < \infty$, and thus, using the fact that $E(\xi_i \xi_j) = 0$, we find that

$$\langle W, f \rangle = \sum \xi_n a_n$$

113

satisfies
$$E(|\langle W, f\rangle|^2) = \sum |a_n|^2 E(|\xi_n|^2) = \|f\|^2.$$

In fact, $\langle W, f\rangle$ is itself a Gaussian random variable with mean zero and variance $\|f\|^2$.

Thus $\langle W, f\rangle$ is, roughly speaking, the value of a random coordinate of f. Unfortunately this does not make W itself an element of H, since it is almost surely not continuous, or even defined everywhere. For example, $f = \sum \xi_n/n$ is almost surely a well-defined element of H, but $\langle W, f\rangle = \sum |x_n|^2/n = \infty$ almost surely.

Smooth vectors. Given an orthonormal basis e_n, we can consider the Sobolev space H_s where f has, so to speak, s derivatives in L^2, or more concretely where
$$\|f\|_{H_s}^2 = \sum n^{2s} |a|_n^2 < \infty.$$

Note that the dual of H_s is H_{-s}. Then W is, almost surely, an element of H_{-s} for s large enough. Indeed, we have for $c \gg 1$,

$$P(|\xi_n| > c) \le \exp(-Ac^2)$$

for some constant A, and thus for any $t > 0$,

$$\sum P(|\xi_n| > n^t) \le \sum \exp(-An^{2t}) < \infty.$$

(To see this, keep in mind that $\sum 1/n^2 = \sum \exp(-2\log n) < \infty$, and n^ϵ goes to infinity faster than $\log n$.)

Thus W is in H_{-t} for any $t > 0$, and hence $B \in H_{1-t}$. Now in n dimensions, having $n/2 + \epsilon$ derivatives in L^2 is enough to certify continuity. Thus B is certainly continuous.

Solving the Dirichlet problem with Brownian motion. Using Brownian motion, we can now give a probabilistic solution to the classical Dirichlet problem on a region $\Omega \subset \mathbb{C}$.

To state the main result, we first study $\partial\Omega$. Let $B(t)$ be Brownian motion in $\mathbb{C}$ started at $B(0) = z$, and let

$$\tau(z) = \inf\{t > 0 \ : \ B(t) \notin \Omega\}.$$

Clearly $\tau(z) > 0$ almost surely if $z \in \Omega$, and $\tau(z) = 0$ if $z \notin \overline{\Omega}$.

In fact, the condition $\tau(z) > 0$ defines a *tail event*, and $P(\tau(z) > 0) = 0$ or 1 for all z. This condition becomes interesting when $z \in \partial\Omega$: in this case, we have z is a *regular* boundary point if

$$P(\tau(z) = 0) = 1.$$

Otherwise z is an *irregular* boundary point, and $\tau(z) > 0$ almost surely. Thus regular bound points behave like they are outside of Ω, and irregular ones like they are inside.

We can now state:

Theorem 10.5 *Let $\Omega \subset \mathbb{C}$ be a bounded open set, and let $F : \partial\Omega \to \mathbb{R}$ be a continuous function. Then*

$$f(z) = E_z(F(B(\tau)))$$

is harmonic in Ω. If every point in $\partial\Omega$ is regular, then f gives the unique continuous, harmonic extension of F.

Proof. Suppose $z \in B = B(z, r) \subset \Omega$. Let σ be the least t such that $B(t) \in \partial B$. We then regard our random walk as the concatenation of a random walk from z to $w = B(\sigma) \in \partial B$, and then a random walk from w to $\partial\Omega$. It follows that $f(z) = E(f(w))$. By symmetry, the point w is uniformly distributed on ∂B. Thus $f(z)$ satisfies the mean value property, and hence it is harmonic.

Now suppose every point $p \in \partial\Omega$ is regular; so $\tau(p) = 0$ almost surely. It follows that, if $z_n \to p \in \partial\Omega$, and $\epsilon > 0$ is given, then

$$P(\tau(z_n) > \epsilon) \to 0.$$

This implies that Brownian paths starting at z_n hit $\partial\Omega$ near p with high probability, when n is large. Since F is continuous (and bounded), it follows that $f(z_n) \to F(p)$. ∎

Regular boundary points. To make this result useful, we need a criterion for regular boundary points. Here are some examples:

1. Let $\Omega = \Delta^*$. Then $p = 0$ is an *irregular* boundary point.

2. If Ω is a Jordan domain, then every point in $\partial\Omega$ is regular. The same holds for finitely connected regions with Jordan boundary.

3. Let $\Omega = \Delta(2) - K$, where K is the middle thirds Cantor set. Then point of K is regular.

4. However, if $K \subset \Delta$ has capacity zero, and $\Omega = \Delta - K$, then no point of K is regular.

5. One can also describe regularity without reference to Brownian motion, using 'barriers'. See [Ah1, Ch. 6, §4.2].

The following simple criterion for regularity is not the most general possible, but it illustrates the idea and handles the case of smooth boundary.

Theorem 10.6 *Suppose there exists a triangle T disjoint from Ω, with p as one of its vertices. Then p is a regular boundary point of Ω.*

Proof. Suppose to the contrary that p is irregular, so $P(\tau(p) > 0) = 1$. Consider a Brownian path $B(t)$ with $B(0) = p$. Let τ_T the least t such that $B(t) \in T$. Then $\tau_T \geq \tau(p)$, and hence $\tau_T > 0$ with probability one. Now choose a finite set of triangles $T_1, \ldots, T_n$, obtained by rotating T about p, such that $\bigcup T_i$ contains a neighborhood of p. By rotation invariance of Brownian motion, $\tau_{T_i} > 0$ with probability one for all i. Thus $\sigma = \min \tau_{T_i} > 0$ almost surely. But this means $B(t) \notin \bigcup T_i$ for all $t < \sigma$. This contradicts the continuity of $B(t)$. ∎

The harmonic image of Brownian motion. (Cf. [MP, Ch. 7], [LeG, §4.3, §7.5].)

We now shift perspective somewhat. Consider Brownian motion $B(t)$ on $\mathbb{R}^d$, starting with $B(0) = 0$, together with a smooth function $f : \mathbb{R}^d \to \mathbb{R}$. Taking the image of a Brownian path, we obtain a new random process

$$C(t) = f(B(t)).$$

If f is *harmonic*, and grows slowly enough that the expected value of $|C(t)|$ is finite, then in fact $C(t)$ is a *martingale*; in particular,

$$E(C(t)) = f(0)$$

is independent of t. (In general, $C(t)$ is a *local martingale*; for its expectation to be finite, we need $f(x)$ to be in the domain of the heat kernel.)

The principle effect of the function $f(x)$ is to modify the speed of Brownian motion. In fact, if we let

$$s = \int_0^t |\nabla f|^2(B(u)) \, du,$$

and define $\widetilde{B}(s) = C(t)$ when s and t are related as above, then $\widetilde{B}(s)$ is *standard Brownian motion* on $\mathbb{R}$. Summing up:

Theorem 10.7 *If $f : \mathbb{R}^d \to \mathbb{R}$ is harmonic, then $f(B(t))$ is time–changed Brownian motion.*

The Itô calculus. The theory of stochastic integrals provides a tool to describe the process $f(B(t))$ for any smooth f. In brief, for a smooth map $f : \mathbb{R}^d \to \mathbb{R}$, we have: *Itô's formula*:

$$f(B(t)) - f(B(0)) = \int_0^t \nabla f(B(u)) \cdot dB(u) + \int_0^t (1/2)\Delta f(B(u)) \, du.$$

This is sometimes written as a *stochastic differential equation*:

$$df(B) = \nabla f \cdot dB + (1/2)\Delta f \, dt.$$

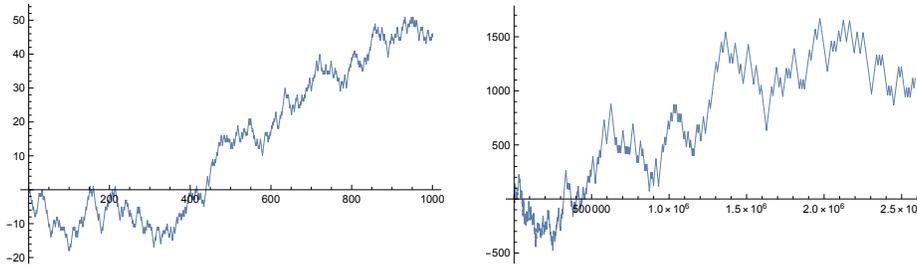A key point here is that the random variable $\nabla f \cdot dB$ has *mean zero*.

Figure 15. Random walks $x(n)$ and $x(n)^2 - n$, with time change.

Example: $x(n)^2$ and $B(t)^2$. To indicate the idea behind this formula, let us first consider $f(x) = x^2$ applied to the discrete random walk $x(n) = \sum_1^n \xi_i$ on $\mathbb{Z}$. Since

$$x(n+1)^2 = x(n)^2 + 2x(n)\xi_{n+1} + 1,$$

117

the process $y(n) = f(x(n)) = x(n)^2$ satisfies

$$
\begin{aligned}
y(n+1) - y(n) &= 2x(n)\xi_{n+1} + 1 \\
&= f'(x(n))\xi_{n+1} + (1/2)f''(x(n)).
\end{aligned}
$$

Thus the second derivative of f introduces a deterministic drift term, apart from which $y(n)$ evolves like a random walk with its steps rescaled by the first derivative of f. The associated new time parameter satisfies

$$
t(n+1) - t(n) = |f'(x(n))|^2 = 4x(n)^2.
$$

Removing the drift term, we obtain a space–time path $(y(n) - n, t(n))$ that behaves, at a large scale, like a standard random walk. For an example, see Figure 15.

Similarly, the process $B(t)^2$ for Brownian motion on $\mathbb{R}$ satisfies:

$$
d(B(t)^2) = B(t)\,dB(t) + \,dt.
$$

Integrating, we find that $C(t) = B(t)^2 - t$ is a martingale; note that its expectation is zero; the associated new time parameter is

$$
s = \int_0^t 4B(u)^2\,du,
$$

and $(C(t), s(t))$ parameterizes an instance of standard Brownian motion in space–time.

Stochastic integration. A critical feature of a stochastic integral such as $\int_0^t g(u)\,dB(u)$ is that $dB(u) = B(u+\delta u) - B(u)$ should be interpreted as the change of $B(u)$ *in the future*. Thus the integral is approximated by sums of the form

$$
\int_0^t g(u)\,dB(u) \approx \sum g(u_i)(B(u_{i+1}) - B(u_i)).
$$

Conformal invariance of Brownian motion. Using the fact that a holomorphic function $f : \mathbb{C} \to \mathbb{C}$ is harmonic, we similarly find that $\widetilde{B}(s) = f(B(t))$ is time–changed Brownian motion, with

$$
s = \int_0^t |f'(B(u))|^2\,du.
$$

For a conformal map between domains, $f : (U, 0) \to (V, 0)$, one should also consider stopping times. Let τ denote the first moment that $B(t)$ exists

U, and σ the corresponding value of s. Let $\widetilde{B}(s) = f(B(t))$ as before. Then $\widetilde{B}(\min(s, \sigma))$ is standard Brownian motion on V, stopped when it reaches the boundary.

Harmonic measure in terms of Brownian motion. Brownian motion provides the following interpretation of harmonic measure.

Theorem 10.8 *The probability that a Brownian path in Ω, starting at z, lands in A when it first reaches $\partial\Omega$, is given by*

$$P_z(B(\tau) \in A) = \omega(z, \Omega, A).$$

Here the stopping time τ is the least t such that $B(t) \in \partial\Omega$.

Since the equilibrium measure, discussed in §5, corresponds to harmonic measure as seen from ∞, we have:

Theorem 10.9 *The equilibrium measure μ on a compact set $K \subset \mathbb{C}$ is the same as the hitting measure for a Brownian path initiated at ∞. In other words,*

$$\mu(A) = P_\infty(B(\tau) \in A),$$

where τ is the least t such that $B(t) \in K$.

This interpretation makes various monotonicity properties of harmonic measure, previously proved with the maximum principle, intuitively obvious. For example, $\omega(z, \Omega, A)$ is an increasing function of Ω, since as Ω grows the stopping time τ increases, and a Brownian path has more chances to reach A.

11 Percolation

In this section we will discuss conformal invariance of the continuum limit of percolation. More precisely, we will present the proof of:

Theorem 11.1 (Smirnov) *Let $Q \subset \mathbb{C}$ be a quadrilateral, isomorphic to the standard rectangle $R(a, b)$. Then the probability $0 < p(Q) < 1$ that a critical percolation cluster connects the a–sides of Q depends only on $\mathrm{mod}(Q) = a/b$.*

We begin by describing percolation on a lattice. We then define the crossing probability $p(Q)$ as a continuum limit, and describe its predicted dependence on Q in terms of its behavior in an equilateral triangle. This dependence motivates the definition of 3 distinct crossing probabilities, defined for z in a topological triangle $T \subset \mathbb{C}$ (a Jordan domain with 3 marked points on its boundary). The main idea in the proof is to show these crossing probabilities can be assembled to form an analytic function. They are thus determined by their boundary values, which are easy to compute.
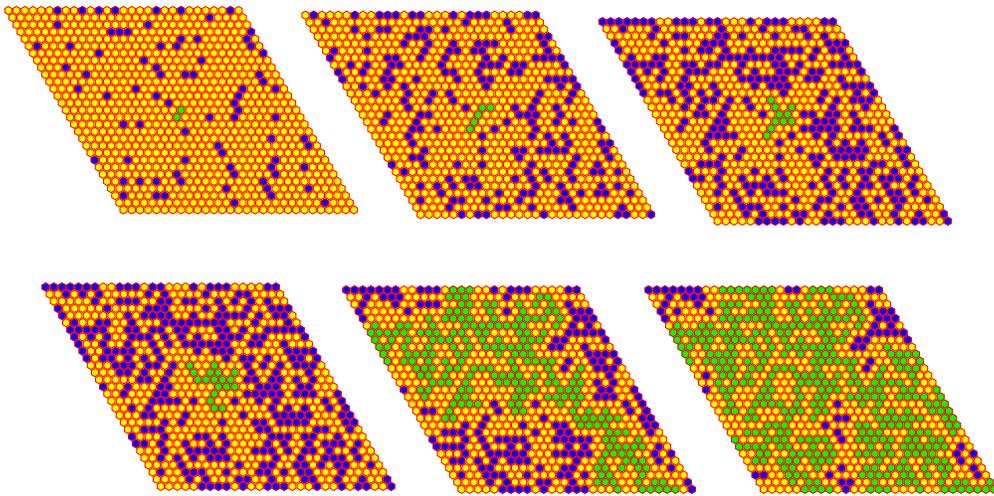

Figure 16. A spreading percolation cluster.

Percolation on a graph. *Percolation* is the study of the behavior of random clusters in a graph G. The concrete example we consider is the following.

Let $\omega = \exp(2\pi i/3) = -1/2 + i\sqrt{3}/2$. Then $\mathbb{Z}[\omega]$ is the *triangular lattice* in $\mathbb{C}$. We make $V(G) = \mathbb{Z}[\omega]$ into the vertices of a graph G by connecting each lattice point z to its six neighbors satisfying $|z - w| = 1$. This graph is the 1–skeleton of a triangulation of the plane.

Now fix a probability p. We can then choose a random subset $S \subset V(G)$ by choosing vertices independently, each with probability p. A *cluster* $S_0 \subset S$ is a maximal edge–connected set.

One of the first question one can consider is, does S contain an infinite cluster S_0? Note that if we remove finitely many vertices from an infinite

cluster, it will break into only finitely many pieces, at least one of which must still be infinite. Thus, by Kolmogorov's 0/1–law, we have:

The probability that S contains an infinite cluster is 0 or 1.

It follows easily that there is a *critical probability p_c* such that the probability is 0 for $p < p_c$, and 1 for $p > p_c$. A basic result is:

Theorem 11.2 *For the triangular lattice, the critical probability is $p_c(\mathbb{Z}[\omega]) = 1/2$.*

This result is not obvious. For example, there could be a range of p where both S and $V(G) - S$ have infinite clusters, or where neither has an infinite cluster. In fact, as we will see below, for $p = p_c$ there *is no infinite cluster!*

An illustration. The behavior of triangular percolation as p passes p_c is illustrated in Figure 16. Each hexagon is centered on a point in the triangular lattice. Initially the central site is green, and the rest are yellow. The picture evolves by a random yellow site turning blue at each step. The blue cluster containing the central site is (re-)colored green. As the density of blue sites passes $1/2$, the green cluster explodes to fill the frame.

Other lattices. Percolation has been studied on many other lattices (such as the square lattice, $\mathbb{Z}[i]$) and other graphs. It is also common to study *bond* percolation, where edges rather than vertices are chosen at random. It is known the $p_c = 1/2$ for *bond* percolation on the square lattice, while for site percolation we have $p_c(\mathbb{Z}^2) \approx 0.59274$; the exact critical value is not known.

The value $1/2$ for the triangular lattice is related to the fact that there are no draws in the game of hex. When two players take turns coloring a hexagonally tiled parallelogram yellow and blue, at the end of the game there is either a blue path joining the horizontal sides, or a yellow path joining the vertical sides. This is not true on a checkerboard: e.g. with the standard alternating coloring, no player has won (or both players have won, depending on how diagonally touching squares are treated). This distinction underlies the difference between site percolation on $\mathbb{Z}[\omega]$ and $\mathbb{Z}[i]$.

It is widely expected that the continuum limit of critical percolation is *robust*, in the sense that it does not depend on the chosen discrete approximation scheme. However Theorem 11.1 has only been proved for the triangular lattice. Thus we will focus on this case.

Critical percolation. We now consider a quadrilateral $Q \subset \mathbb{C}$ with cyclically ordered marked points (q_1, q_2, q_3, q_4). The a–sides of Q are the pair of arcs $\alpha = [q_1, q_2]$ and $\alpha' = [q_3, q_4]$ in ∂Q.

Let $h \to 0$ be a sequence of mesh sizes. As usual, we can then approximate Q by a subgraph of G with vertices

$$Q_h = Q \cap h\mathbb{Z}[\omega].$$

By approximating the a–sides of Q by arcs in ∂Q_h, we can regard Q_h as a combinatorial quadrilateral. Setting $p = p_c = 1/2$, we define

$$p(Q_h) = P(\text{there exists a cluster } S_0 \text{ connecting the } a\text{–sides of } Q_h).$$

Implicit in the statement of Theorem 11.1 is the fact that the limit

$$p(Q) = \lim_{h \to 0} p(Q_h)$$

exists.

The value of $p(Q)$ has a remarkably simple description. Given conformal invariance, it suffices to compute the value when Q is an equilateral triangle $T_0 \subset \mathbb{C}$.

Theorem 11.3 *Let $T_0 \subset \mathbb{C}$ be an equilateral triangle with vertices ABC. Let X be a point on the edge AB. Then the probability of a percolation cluster joining AX to the opposite edge BC is given by $|AX|/|BC|$.*

Using hypergeometric functions, one can then give an explicit formula for $p(Q)$ in terms of $\mathrm{mod}(Q)$. This complicated formula was first conjectured by Cardy. It was Carleson who observed that the formula becomes very simple for an equilateral triangle. The simplification is the first indication that percolation is especially well adapted to the triangular lattice.

Barycentric coordinates on a triangle. To formulate a still more precise statement, it is useful to normalize the equilateral triangle T_0 so its vertices are $(1, \omega, \omega^2)$. Let (b_0, b_1, b_2) be *barycentric coordinates* on T. These are real–valued functions satisfying

$$b_0 + b_1 + b_2 = 1 \quad \text{and} \quad z = b_0 + \omega b_1 + \omega^2 b_2,$$

for all $z \in T$. Note that b_i vanishes on the edge of T opposite the vertex ω^i. It is useful to think of the indices $0, 1, 2$ as lying in $\mathbb{Z}/3$. We note that

$$b_i(z) = b_{i+1}(\omega z),$$

122

since

$$\omega z = \omega b_0(z) + \omega^2 b_1(z) + b_2(z).$$

Note that $b_i(z)$ is a *harmonic function* of z for each i; for example, we have

$$b_0(z) = (2\operatorname{Re}(z) + 1)/3.$$

Conformal triangles. The analogue of a quadrilateral, with only 3 vertices, is a *conformal triangle*, by which we mean a Jordan domain $T \subset \mathbb{C}$ together with a positively ordered triply of distinct points (t_0, t_1, t_2) in ∂T. There is a unique Riemann mapping $f : T \to T_0$ such that $f(t_i) = \omega^i$: a conformal triangle has no moduli. We refer to the arc $[t_{i+1}, t_{i+2}] \subset \partial T$ as the *ith side* of T; equivalently, it is the edge *opposite T_i*.

Any conformal triangle carries barycentric coordinates, pulled back from those on T_0.

Theorem 11.4 *Barycentric coordinates on T are characterized, among all C^1 functions $b_i : T \to [0, 1]$, by the following properties for $i = 0, 1, 2$:*

1. *We have $\omega \overline{\partial} b_i = \overline{\partial} b_{i+1}$; and*

2. *We have $b_i(t_i) = 1$ and $b_i = 0$ on the opposite edge of T.*

Note. If one sets $\nabla f = (df/dx) + i(df/dy)$, then second condition can be written $\omega \nabla b_i = \nabla b_{i+1}$. In other words, at every point z the 3 vectors $\nabla b_i(z)$ point to the vertices of an equilateral triangle. One should compare this condition to the usual Cauchy–Riemann equations for $f(z) = a_0 + ia_1$, which say that $i\overline{\partial} a_0 = \overline{\partial} a_1$.

Proof. The $\overline{\partial}$ conditions imply that

$$\overline{\partial}(b_0 + b_1 + b_2) = (\overline{\partial} b_0)(1 + \omega + \omega^2) = 0;$$

thus the real function $b_0 + b_1 + b_2$ is holomorphic, and hence constant; and in view of the boundary conditions, we have:

$$b_0(z) + b_1(z) + b_2(z) = 1 \tag{11.1}$$

for all z.

Similarly, if we let

$$f(z) = b_0(z) + \omega b_1(z) + \omega^2 b_2(z),$$

123

then
$$\overline{\partial} f = \overline{\partial} b_0 (1 + \omega^2 + \omega) = 0,$$

so $f(z)$ is holomorphic. Now the boundary conditions together with equation (11.1) imply that $f(t_i) = \omega^i$ on the vertices of T. Moreover, on 0th side of T, we have $b_0 = 0$ and hence

$$f(z) = \omega b_1 + \omega^2 b_2$$

with $b_1 + b_2 = 1$ and $b_1, b_2 \geq 0$. Thus f maps the 0th edge of T to the 0th edge of T_0. The same holds true for the other edges. It follows that as z moves once around ∂T, $f(z)$ moves once around any given point w in the interior of T_0. Thus by the argument principle, $f : T \to T_0$ is a conformal homeomorphism, and in fact f is the standard isomorphism since $f(t_i) = \omega^i$. ∎

Here is a variant of the result above, which follows by the same argument:

Theorem 11.5 *Barycentric coordinates on T are characterized by their required boundary values and the property that $b_0 + b_1 + b_2$ and $b_0 + \omega b_1 + \omega^2 b_2$ are holomorphic functions.*

The advantage of this statement is that uniform limits of holomorphic functions are holomorphic.

Percolation in a triangle. We can now formulate a result that combines Theorems 11.1 and 11.3.

As before, we can now study critical percolation on the lattice approximations
$$T_h = T \cap h\mathbb{Z}[\omega].$$

We think of T_h as the vertices of a finite triangular graph G_h, which is just a triangulation of a topological disk. We will assume that the marked points t_i are actually vertices in ∂G_h. These points divide the boundary into three *sides*; we refer to the arc

$$\alpha_i^h = [t_{i+1}, t_{i+2}] \subset \partial G_h$$

as *side i* of G_h. It is the side opposite t_i.

As usual we study critical percolation with $p = 1/2$ on G_h, and let $S \subset V(G)$ be a random set of vertices. Given a point z in the *center* of one of the triangles of G_h, we let

$B_i^h(z) = P(\text{there exists a simple path in } S \text{ separating } z \text{ from side } i).$

We emphasize that this path is *simple*: it cannot visit a vertex in S twice. For example, in Figure 17, the vertices in S are colored blue, those not in S are yellow, and the triangles separated from side 0 are shaded light blue. Note that the triangle below w is *not* separated from side 0 by a *simple* path, even though there is a blue barrier between it and side 0.

We may now state:

Theorem 11.6 *For any conformal triangle T, the functions $B_i^h(z)$ converge uniformly, as $h \to 0$, to barycentric coordinates on T.*

For $z \in \partial T_h$, $B_0^h(z)$ is the same as the probability of a percolation cluster separating $[t_0, z]$ from $[t_1, t_2]$. Thus Theorem 11.6 implies Theorems 11.1 and 11.3.

The Riemann mapping theorem. Note that the theorem above implies that *percolation computes the Riemann mapping* $f : T \to T_0$, since $f(z) = \sum \omega^i b_i(z)$.

Outline of th proof. The rest of this section is devoted to sketching the proof of Theorem 11.6. There are three main steps.

I. *Limits exist.* We will show that the functions B_i^h have a uniform modulus of continuity, and hence along a subsequence we have uniform convergence to a continuous limits $B_i = \lim B_i^h$, $i = 0, 1, 2$.

One can compare this step to the Lipschitz bounds for combinatorial harmonic functions (Corollary 9.5).

II. *Triangular Cauchy–Riemann equations.* We will establish a combinatorial version of the system of equations

$$\omega \overline{\partial} B_i^h = \nabla B_{i+1}^h.$$

This step is the key to the proof. It takes the form of an exact equality between certain probabilities, and it is here that the triangular lattice plays a special role.

Using I and II, one can show that the functions $B_0^h + B_1^h + B_2^h$ and $B_0^h + \omega B_1^h + \omega^2 B_2^h$ nearly integrate to zero over loops in T. It follows that their uniform limits are *holomorphic*.
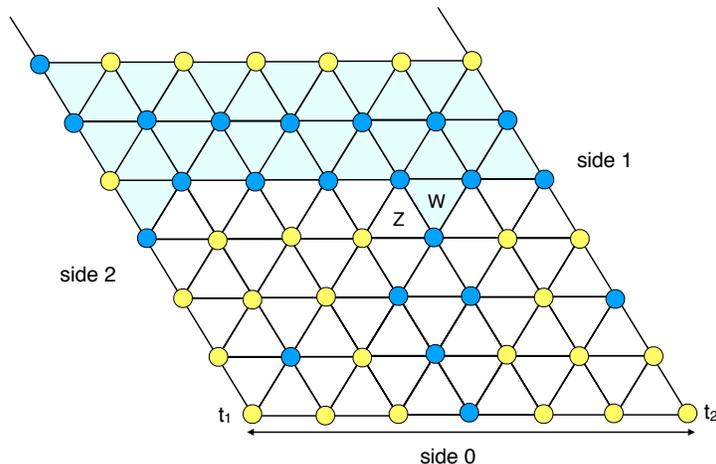
125

Figure 17. The shaded region is separated from side 1 by a blue simple curve.

III. *Boundary conditions.* Finally one check that the limiting functions $B_i(z)$ satisfy the same boundary conditions as $b_i(z)$, and hence they are equal, by Theorem 11.5.

This last step is straightforward. We will concentrate on the main ideas for steps I and II.

I: Continuity and RSW theory. The first step relies on the following result, due to Russo, Seymour and Welsh, which also indicates the state of the art in percolation prior to Smirnov's work.

Theorem 11.7 *The crossing probabilities for a standard rectangle $Q = R(a, b)$ satisfy*

$$0 < \liminf p(Q_h) \leq \limsup p(Q_h) < 1,$$

the limits taken as the mesh $h \to 0$. A similar statement holds for clusters crossing the standard round annulus $A(R)$.

Here a cluster *crosses* $A(R)$ if it joins its two boundary components.

This statement is not particular to the triangular lattice. We will sketch two ideas behind the proof: positive correlation of monotone events, and bootstrapping. For more details, see [BR, Ch. 3].

Monotone events. The probability space for a sequence of n independent, fair coin flips is a Boolean cube $Q_n = \{0, 1\}^n$ with 2^n points, each with equal

126

weight. Equivalently, we can regard $Q_n = \mathcal{P}(E)$ as the power set of a set E with $|E| = n$.

We will be using $x \in Q_n$ to parameterize the choice of a random set of edges $E = E(G)$ in a graph, where the edge e_i is included if $x_i = 1$ and otherwise not.

An *event* is a subset of $A \subset Q_n$; its probability is $P(A) = |A|/|B_n| = 2^{-n}|A|$. An event is *monotone* (increasing) if whenever $x \in A$ and $y_i \geq x_i$ for all i, we have $y_i \in A$. In the language of sets, whenever $X \subset E$ is in A and $Y \supset X$, we have $Y \in A$.

Monotone events occur frequently when we consider connectivity. E.g. the event that a set of edges $X \subset E$ joins two vertices $p, q \in V(G)$ is monotone, since adding more edges can only improve matters. In logic, a monotone event is one that can be written as $\bigvee w_i$ where each w_i is a conjunction of a subset of the logical variables $t_1, \ldots, t_n$, where their negations $\bar{t}_i$ do *not* appear.

(Such a logical expressions are easy to simplify: their minimal expression in conjunctive form is obtained by throwing out w_i whenever $w_j \implies w_i$ for some j. The general problem of log simplification is NP hard.)

Positive correlation. Two events A and B are *independent* if $P(AB) = P(A)P(B)$. They are *positively correlated* if $P(AB) \geq P(A)P(B)$, or equivalently if
$$P(A|B) \geq P(A).$$
(This is the same as $P(B|A) \geq P(B)$.

It is intuitively obvious that a pair of monotone events are positively correlated. This is a very useful observation, due to Harris.

Theorem 11.8 *Any two monotone increasing events are positively correlated.*

Proof. We will use 'divide and conquer' as is common in the study of logic. Namely we partition Q_n into the space where $e_n = 0$ and the space where $e_n = 1$, and write the induced partition of subsets by $A = A_0 \sqcup A_1$. We then proceed to prove the theorem by induction on n, using the fact that A_i and B_i depend on fewer variables.

The key inequality, based on monotonicity in the last variable, is:
$$(P(A_1) - P(A_0))(P(B_1) - P(B_0)) \geq 0,$$

or in other words

$$P(A_0)P(B_0) + P(A_1)P(B_1) \geq P(A_0)P(B_1) + P(A_1)P(B_0). \qquad (11.2)$$

On the other hand, induction gives

$$P(A_iB_i) \geq 2P(A_i)P(B_i)$$

for $i = 1, 2$, where we have rescaled the measure on $(Q_n)_i$ by a factor of two to make it a probability measure. This shows that

$$P(AB) = P(A_0B_0) + P(A_1B_1)$$

is bounded below by the sum of the right and left hand sides of (11.2), which is just

$$(P(A_0) + P(A_1))(P(B_0) + P(B_1)) = P(A)P(B).$$

■

Bootstrapping. We now return to critical percolation on the triangular lattice with lattice $\mathbb{Z}[\omega]$ and graph G. We consider for concreteness integers $a, b > 0$ and *finite* 'rectangles' (actually parallelograms) $L(a, b)$, with vertices at $0, a, \omega b$ and $a + \omega b$. The proof of Theorem 11.7 proceeds as follows.

1. First, one observes that for a rhombus, we have $p(L(a, a)) = 1/2$. This observation is related to the game of hex; it uses the symmetry of the rhombus and the fact that $p_c = 1/2$. Let A be the event that a random set of vertices S join the horizontal sides of the rhombus, and B the event that the remaining vertices join the vertical sides. It is easy to see that $P(A) = P(B)$ and that exactly one of these events must occur; thus both have probability $1/2$.

2. Second, one shows there is a $q > 0$ such that

$$q < p(L(a, 2a)) < 1 - q$$

for all a. This step is the key to the proof. It is enough to establish this step with 2 replaced by any constant > 1.

3. Next, one shows $q^{2k} < p(L(a, ka))$ for all k. (The upper bound $p(L(a, ka)) < 1 - q$ is automatic.

   This bound is obtained by bootstrapping. We cover the region between the top and bottom of $L(a, ka)$ by copies L_i of $L(a, 2a)$, $i = 1, 2, \ldots, k - 1$. This can be done in such a way that $L_i \cap L_{i-1} = M_i \cong L(a, a)$ is a rhombus.

   By the monotonicity theorem, crossing events are positively correlated. Thus the probability that there are vertical crossings in all the L_i, and horizontal crossings in all the M_i, is at least q^{2k}. But all these crossings taken together give a vertical crossing in $L(a, ka)$.

4. The upper and lower bounds for an annulus are obtained similarly, using overlapping parallelograms.

No infinite clusters. Theorem 11.7 easily yields:

Corollary 11.9 *For critical percolation on $\mathbb{Z}[\omega]$, the probability of an infinite cluster is zero.*

Proof. Let $A_n \subset \mathbb{Z}[\omega]$ be a sequence of disjoint combinatorial annuli, approximating the regions $10^n < |z| < 2 \cdot 10^n$. Let B_n be the event that there is no cluster crossing A_n. The events B_n are independent, and $P(B_n) > q > 1$ by Theorem 11.7. Thus $P(\limsup B_n) = 1$. But any infinite cluster must cross all but finitely many B_n, so the probability that such a cluster exists is zero. ∎

The same idea lies behind:

Theorem 11.10 *The functions $B_i^h(z)$ are uniformly Hölder continuous. That is, there exists $C, \alpha > 0$ such that*

$$|B_i^h(z) - B_i^h(w)| \leq C|z - w|^\alpha.$$

Proof. Let $S \subset V(T_h)$ be a random set of vertices. Suppose there is an annulus $A \subset T_h$, not crossed by S, that encloses z and w. Then one can picture z and w as living in a bubble B, whose boundary is made up of sites outside of S. For S to separate one point of the bubble from the vertex t_i, it must separate all points in the bubble from t_i.

Let $q(z, w)$ be the probability that z and w do not lie in a bubble, and let $r = |z - w|$. Then $|B_i^h(z) - B_i^h(w)| \le q(z, w)$. Now we can enclose $\{z, w\}$ in approximately $N = \log(1/r)$ disjoint annuli A_n, each of definite modulus. If z and w are not in a bubble, then S must cross every A_n. The probability S crosses A_n is $\le q < 1$, and hence

$$|B_i^h(z) - B_i^h(w)| = O(q^N) = O(r^\alpha),$$

where $\alpha = \log(1/q)$. (Note that when q is close to 1, α is close to zero.) $\blacksquare$

II. The combinatorial Cauchy–Riemann equations. We now turn to the study of the combinatorial gradient of $B_i(z)$. Recall that z ranges through the centers of the triangles in the graph G_h with $V(G_h) = T_h$. Every center z is adjacent to 3 others, each of the form

$$w_j = z + \eta \omega^j$$

for $j = 0, 1, 2$, where $\eta = \pm ih/\sqrt{3}$ and the sign depends on the triangle, which can point up or down.

Given two adjacent centers z and w, we let

$$D_i(z, w) = P(w \text{ is separated from side } i \text{ of } T_h, \text{ but } z \text{ is not}).$$

As usual we require the separation to come from a *simple* path in $S \subset T_h$, with endpoints on the of T_h sides adjacent to t_i. We then have

$$B_i(z) - B_i(w) = D_i(w, z) - D_i(z, w).$$

The next result is the analogue of the triangular Cauchy–Riemann equation, $\omega \bar{\partial} b_i = \bar{\partial} b_{i+1}$. It is an *equality* of probabilities that holds in each lattice approximation T_h to T.

Theorem 11.11 *For all z and all i, j, we have*

$$D_i(z, w_j) = D_{i+1}(z, w_{j+1}).$$

Proof. The argument is illustrated in Figure 17, for $i = 0$.

Let (v_0, v_1, v_2) label the vertices of the triangle containing z, so that v_j is opposite w_j. In the picture, $w = w_j$ is separated from side 0 of T, but z is not. Thus this configuration S contributes to $D_0(z, w_j)$.

Note that the separation properties of this configuration are witnessed by disjoint simple paths γ_{j+k}, connecting vertex v_{j+k} to side $k = 0, 1, 2$. The paths $(\gamma_j, \gamma_{j+1}, \gamma_{j+2})$ are colored yellow, blue and blue respectively. The two blue paths form the simple arc separating w_j from side 0.

Now there may be many paths of each type. Choose the yellow and blue paths, γ_j and γ_{j+1}, so they are as close together as possible. They then form the boundary of a region Ω. Now flip the colors in Ω. This has no effect the probability that the third blue path γ_{j+2} exists. But it changes γ_{j+1} to yellow. Since this path leads to side 1, and the other two separate w_{j+1} from side 1, these new configuration contributes to $D_1(z, w_{j+1})$. ∎

Sketch of the proof of Theorem 11.6. By Theorem 11.10, along a subsequence $h = h_n arrow 0$ we have continuous limits $B_i(z) = \lim B_i^h(z)$. Using Theorem 11.11, one can show that the functions

$$s(z) = B_0(z) + B_1(z) + B_2(z) \quad \text{and} \quad f(z) = B_0(z) + \omega B_1(z) + \omega B_2(z)$$

integrate to zero around enough closed loops that they are holomorphic. Thus the functions $B_i(z)$ are harmonic. Using Theorem 11.5 and their boundary conditions, we conclude that $B_i(z) = b_i(z)$ are in fact barycentric coordinates on T. ∎

12 Problems

1. Prove or disprove: if $f : \Delta \to \mathbb{C}$ is an analytic function, and $f'(z)$ is never zero, then f is injective.

2. Let $f : \Delta \to \mathbb{C}$ be a local homeomorphism that is not injective. Prove that there is no sequence of injective continuous functions $f_n : \Delta \to \mathbb{C}$ converging uniformly to f.

3. Construct a surjective analytic map $f : \Delta \to \mathbb{C}$. (Hint: write f as a composition of easily understood maps.)

   Can one find such a surjective map with $f'(z) \neq 0$ for all $z \in \Delta$?

4. Show there is no proper analytic map $f : \Delta \to \mathbb{C}$. Here *proper* means $f^{-1}(K)$ is compact whenever $K \subset \mathbb{C}$ is compact. (Hint: such a map has finitely many zeros, $a_1, \ldots, a_n \in \Delta$; consider $f(z) \prod_{i=1}^n (1 - \bar{a}_i z)/(z - a_i)$.)

5. Find all analytic functions $f : \Delta \to \Delta$ such that $f(0) = 0$ and $f(1/2) = 1/4$. Your solution should be of the form $f(z) = P(z, g(z))$, where $g : \Delta \to \overline{\Delta}$ is an arbitrary analytic function.

6. Let $f : \Delta^* \to \Delta^*$ be an analytic map that induces the homomorphism $n \mapsto \deg(f) \cdot n$ on $\pi_1(\Delta^*) \cong \mathbb{Z}$.

   (i) Show that $\deg(f) \geq 0$.

   (ii) Show that if $\deg(f) > 0$, then $|f(z)| \leq |z|^{\deg(f)}$ for all $z \in \Delta^*$.

7. For $0 < \alpha < 2\pi$, let P_α be the pie slice of the punctured unit disk defined by

$$P_\alpha = \{z \in \mathbb{C} : 0 < |z| < 1 \quad \text{and} \quad 0 < \arg(z) < \alpha\}.$$

   Give an explicit Riemann mapping $f : P_\alpha \to \Delta$.

8. (Continuation.) Compute the hyperbolic metric on P_α, and verify that the inclusion $P_\alpha \subset \Delta$ is a contraction for the hyperbolic metric on Δ.

9. Let $S_+ \subset \mathbb{C}$ be the region defined by $\mathrm{Re}(z) > 0$ and $0 < \mathrm{Im}(z) < \pi$. Find the hyperbolic metric on S_+.

   Verify that the inclusion $S_+ \to S$ is a contraction for the hyperbolic metric on $S = \{z : 0 < \mathrm{Im}(z) < \pi\}$.

10. For $0 < \alpha < 1$, find the hyperbolic metric on the region

$$H_\alpha = \{z : 0 < \arg(z) < \pi\alpha\}.$$

   Verify that the inclusion $H_\alpha \subset \mathbb{H}$ is a contraction for the hyperbolic metric on $\mathbb{H}$.

11. Let $B \subset \mathbb{C}$ be the closed unit ball centered at $z = i$.

   (a) Find an explicit conformal map $f : \mathbb{H} - B \to \mathbb{H}$.

   (b) Find an explicit conformal map $f : \Delta - [0, 1] \to \Delta$.

12. Find a region $U \subset \mathbb{C}$ such that $\cos : U \to \mathbb{H}$ is a conformal isomorphism.

13. Let $\alpha = -iy$ be a *purely imaginary* complex number with $y > 0$. Let $f : \mathbb{H} \to \mathbb{C}$ be the map defined by

$$f(z) = z^\alpha = \exp(\alpha \log z),$$

using the usual single–valued branch of the logarithm in $\mathbb{H}$.

(i) Show that $f(\lambda z) = f(z)$ where $\lambda = \exp(2\pi/y)$.

(ii) Show that f defines a covering map

$$f : \mathbb{H} \to A(R) = \{z \ : \ 1 < |z| < R\},$$

where $R = \exp(\pi y)$.

(iii) Show that f sends the positive imaginary axis to the circle $|z| = \sqrt{R}$, and conclude that this circle is a geodesic of length $2\pi/y$ in the hyperbolic metric on $A(R)$.

14. Show that the hyperbolic metric on $A(R)$ is given by

$$\rho_{A(R)} = \frac{\pi/\log R}{\sin(\pi \log |z|/\log R)} \frac{|dz|}{|z|}. \tag{12.1}$$

(You may use the previous exercise.)

15. Let $U \subset \mathbb{C}$ be a simply–connected region, other than $\mathbb{C}$ itself, and let $p \in U$. Suppose in the proof of the Riemann mapping theorem we had defined

$$\mathcal{F} = \{f : (U, p) \to (\Delta, 0) \ : \ f \text{ is analytic and } f'(0) \geq 0.\}.$$

Prove that (i) there exists an $f \in \mathcal{F}$ that maximizes $f'(0)$ and (ii) that f is in fact the unique Riemann mapping from (U, p) to $(\Delta, 0)$ with $f'(p) > 0$. (Hint: apply the Schwarz Lemma to $f \circ F^{-1}$, where $F \in \mathcal{F}$ is the Riemann map.)

16. Let $J \subset \mathbb{C}$ be a Jordan curve, i.e. a compact set homeomorphic to S^1. Show there exists a positive function $\delta(r)$, defined for $r > 0$, such that for all $x, y \in J$,

$$|x - y| < \delta(r) \implies \operatorname{diam} J_{xy} < r,$$

where J_{xy} is the smallest subarc of J containing x and y.

17. Let $f : S^1 \to \mathbb{C}$ be a continuous function. Show that the image $K = f(S^1)$ is a locally connected topological space. (This means the topology on K has a basis of connected open sets.)

18. Let $z_n \in \mathbb{H}$ satisfy $z_n \to 0$. Recall that $\gamma = i\mathbb{R}_+$ is a geodesic in the hyperbolic metric.

    Show that following two conditions are equivalent:

    (i) $\sup d_{\mathbb{H}}(z_n, \gamma) < \infty$ in the hyperbolic metric on $\mathbb{H}$;

    (ii) $\inf \arg(z_n) > 0$ and $\sup \arg(z_n) < \pi$, where $\arg(z) \in (0, \pi)$.

    In either case we say $z_n \to 0$ *radially*.

19. (Continuation.) Let $f : \mathbb{H} \to U$ be a Riemann mapping to a bounded domain $U \subset \mathbb{C}$. Suppose that $\lim_{y \to 0+} f(iy) = w$.

    (i) Show that $w \in \partial U$.

    (ii) Show that $f(z_n) \to w$ whenever $z_n \to 0$ radially.

    (Hint: show that $\sup d_U(f(z_n), f(i|z_n|)) < \infty$ in the hyperbolic metric on U, and then use the comparison between ρ_U and the $1/d$ metric.)

20. Show that for every $r > 0$ there exists an injective analytic map $f : \Delta \to \widehat{\mathbb{C}}$ with $f(0) = 0$ and $f'(0) = 1$, such that $r \notin f(\Delta)$. Why does this not contradict the Koebe $1/4$ theorem?

21. Let $u : \mathbb{H} \to \mathbb{R}_+$ be a positive harmonic function. Show that u is distance non–increasing from the hyperbolic metric on $\mathbb{H}$ to the hyperbolic metric on $\mathbb{R}_+$, defined by $\rho_{\mathbb{R}_+} = |dx|/|x|$. That is, show that

    $$d_{\mathbb{H}}(z, w) \geq d_{\mathbb{R}_+}(u(z), u(w)),$$

    for all $z, w \in \mathbb{H}$. (Hint: first show there is an analytic function $f : \mathbb{H} \to \mathbb{H}$ such that $u(z) = \mathrm{Im}\, f(z)$.)

22. Prove that any positive harmonic function on $\mathbb{C}$ is constant.

23. Find a bounded harmonic function $\phi(z)$ on the unit disk Δ such that for $z \in S^1$,

    $$\lim_{r \to 1-} \phi(rz) = 1$$

    if $\mathrm{Im}(z) > 0$, and the limit is 0 if $\mathrm{Im}(z) < 0$.

24. Let f be a compactly supported smooth function on $\mathbb{C} \cong \mathbb{R}^2$, with coordinates $z = r\exp(i\theta)$. Show that

$$f(0) = -\frac{1}{2\pi} \int_{\mathbb{C}} \frac{df}{dr} \, dr \, d\theta.$$

25. Let $\phi(z) = \log|z|$. Prove that for any compactly supported function f on $\mathbb{C}$, we have

$$\int_{\mathbb{C}} \phi(\Delta f) = Cf(0),$$

and compute the value of C. This is the precise meaning of the assertion that $\Delta\phi = C\delta_0$ as a distribution. (Hint: approximate by an integral over $\Omega_r = \{z \,:\, r < |z|\}$, and apply Stokes' theorem and the previous exercise.)

26. Let $\phi(z) = \operatorname{Re}(1/z)$. This function is harmonic on $\mathbb{C}^*$. What is the flux of $\nabla\phi$ through S^1?

The distribution $\Lambda = \Delta\phi$ is defined by

$$\Lambda(f) = \int_{\mathbb{C}} (\Delta f)\phi$$

for every smooth, compact supported function f on $\mathbb{C}$. What is $\Lambda(f)$ in terms of the values of f and its derivatives at $z = 0$?

27. Let $U \subset \mathbb{C}$ be a bounded, simply–connected region such that ∂U is locally connected. Prove that the Dirichlet problem (with continuous boundary data) has a solution on U.

28. Compute the Dirichlet energy

$$D(\phi) = \int_{\Delta} d\phi \wedge *d\phi$$

of the harmonic function $\phi(x + iy) = x^3 - 3xy^2$ on the unit disk. (Hint: no integration is necessary.)

29. Determine the Poisson kernel for $\mathbb{H}$. That is, find a function harmonic function $P(x + iy) = P_y(x)$ such that the harmonic extension of $\Phi \in C(\widehat{\mathbb{R}})$ to $\mathbb{H}$ is given by

$$\phi(x + iy) = \frac{1}{\pi} \int_{\mathbb{R}} \Phi(t) P_y(x - t) \, dt.$$

What analytic function $f(z)$ satisfies $P(z) = \operatorname{Im} f(z)$?

30. Let U be a bounded convex region, $z \in U$, and let $A \subset \partial U$ be an arc with endpoints a, b, and let θ be the angle at z of the triangle with vertices (a, b, z). Prove that

$$\omega(z, U, A) \le \theta/\pi.$$

31. Let $\Omega = \{z : 0 < \arg(z) < \pi/2\}$ be the first quadrant. How does the harmonic measure

$$h(t) = \omega(t(1 + i), \Omega, [0, 1])$$

behave as $t \to +\infty$? (An estimate such as $h(t) \sim Ct^\alpha)$ or $h(t) \sim Ce^{-t}$ is sufficient.)

32. Let $V \subset \mathbb{Z}^d$ be a finite set, and let ∂V denote the points $y \in \mathbb{Z}^d - V$ such that $|x - y| = 1$ for some $x \in V$. Let $\overline{V} = V \cup \partial V$.

Define the discrete Laplacian $A : \mathbb{R}^{\overline{V}} \to \mathbb{R}^V$ by

$$Af(x) = (2d)^{-1} \sum_{|y-x|=1} f(y) - f(x)$$

where the sum is over the $(2d)$ lattice points y adjacent to x.

(i) Prove that harmonic functions on $\overline{V}$ satisfy the maximum principle. That is, if $Af = 0$, then $\max_{\partial V} f(v) = \max_V f(v)$.

(ii) Prove the Dirichlet problem has a solution for V. That is, for any function $F : \partial V \to \mathbb{R}$, there exists a harmonic function $f : \overline{V} \to \mathbb{R}$ such that $f|\partial V = F$. (Hint: use the linear map A, and (i), to calculate the dimension of the space of harmonic functions.)

33. Consider the metric $\rho = \rho(z)|dz|$ on $\widehat{\mathbb{C}}$, where $\rho(z) = 1$ for $|z| \le 1$ and $\rho(z) = 2/(1+|z|^2)$ for $|z| \ge 1$. If we regard curvature of ρ as a measure, what is its restriction to the unit circle, $|z| = 1$?

34. Fix $R > 1$, and let $U = \{z : 0 < |z| < R\}$ and let $V = f(U)$, where $f(z) = 1/z$. Let $A = U \cap V$.

(i) Show that the hyperbolic metrics ρ_U and ρ_V agree on S^1.

(ii) Let ξ be the metric on A given by ρ_U when $|z| \geq 1$ and ρ_V when $|z| \leq 1$. Show that ξ is an ultrahyperbolic metric.

(iii) Check that $\xi \leq \rho_A$ throughout A, as predicted by Ahlfors' Schwarz Lemma. You may use equation (12.1).

35. Prove Hadamard's inequality: if M is an $n \times n$ complex matrix, with columns $v_i \in \mathbb{C}^n$, then
$$|\det M| \leq \prod_1^n \|v_i\|.$$

(Hint: reduce to the case where $\|v_i\| = 1$ for all i, and consider the self–adjoint matrix $A = M^*M$; then $\mathrm{Tr}(A) = n$ and it suffices to show $|\det A| \leq 1$.)

36. Show that for any monic polynomial $P_d(z)$ of degree d, $\max_{z \in \overline{\Delta}} |P_d(z)| \geq 1$, and equality holds if and only if $P_d(z) = z^d$. (Hint: first show the average of $P_d(z)/z^d$ over S^1 is 1.)

37. Show that for $(a_1, \dots, a_n)$ the nth roots of unity $(a_k = \exp(2\pi i k/n))$, we have
$$D_n(a) = \prod_{i<j} |a_i - a_j| = (\sqrt{n})^n.$$

38. (Continuation.) Show that $D_n(a) \leq (\sqrt{n})^n$ for every vector $a = (a_1, \dots, a_n)$ with $|a_i| \leq 1$. (Hint: use the Vandermonde determinant, Hadamard's inequality and that fact that $\|a\| \leq \sqrt{n}$.)

Conclude that $\max D_n(a) = (\sqrt{n})^n$, where the maximum is taken over all n-tuples with $a_i \in \overline{\Delta}$.

39. Find the equilibrium measure on $K = [-2, 2]$.

40. Let $P_n : \mathbb{C} \to \mathbb{C}$ be a monic polynomial of degree n, and let $K \subset \mathbb{C}$ be a compact set. Prove that
$$\mathrm{cap}(P_n^{-1}(K)) = \mathrm{cap}(K)^{1/n}.$$

41. Find the capacity of the n–pointed star $K_n \subset \mathbb{C}$ formed by the radii of the unit disk joining $z = 0$ to the nth roots of unity.

42. (Continuation.) Find an explicit conformal mapping from the outside of the unit disk to $\mathbb{C} - K_n$.

43. Show that the standard middle–thirds Cantor set $K \subset [0,1]$ has positive capacity. (Hint: one can write $K = \bigcap K_n$ where K_n consists of 2^n intervals, each of length 3^{-n}. Use the endpoints of the intervals forming K_n to bound the transfinite diameter of K from below.)

44. Show that there exists a closed, uncountable set $K \subset [0,1]$ of zero capacity. (Hint: construct a Cantor set of Hausdorff dimension zero, and study its transfinite diameter.)

45. Show that the capacity of an arc $A \subset S^1$ of length L is given by $\mathrm{cap}(A) = \sin(L/4)$. (Hint: there is $g \in \mathrm{Aut}(\widehat{\mathbb{C}})$ such that $g(A) = [-2,2]$.)

46. Show that if $A \subset \mathbb{R}$ is a compact set, of linear measure L, then $\mathrm{cap}(A) \geq \mathrm{cap}([0,L]) = L/4$.

47. Compute the limit, as $R \to \infty$, of the capacity of $K_R = [0,1] \cup [R, R+1]$.

48. Let $K(P)$ be the filled Julia set of the polynomial $P(z) = z^2 - 1$. The Riemann mapping
$$f : (\mathbb{C} - \overline{\Delta}) \to (\mathbb{C} - K(P)),$$
normalized so that $f(z) = z + \sum_0^\infty b_n/z^n$, satisfies
$$f(z^2) = P(f(z)).$$
Using this functional equation, compute b_n for $n = 0, 1, 2, 3$. (Hint: first show f is an odd function.)

49. Let $m(z) = m(x+iy) = \max(|x|, |y|)$ and consider the 'square annulus' $A = \{z : 1 < m(z) < 10\}$. (i) Using the Euclidean metric, give upper and lower bounds on the modulus $\mathrm{mod}(A)$.

(ii) Using a different metric, improve either your upper or your lower bound.

50. Let $Q \subset \mathbb{C}$ be a quadrilateral, with labeled boundary points (q_1, q_2, q_3, q_4). (i) Prove that there is a homeomorphism $f : \overline{Q} \to \overline{Q}$, analytic on Q, such that $f(q_i) = q_{i+2}$, where the indices are taken $\mathrm{mod}\,4$. (ii) Prove that, in general, there is no such map f such that $f(p_i) = p_{i+1}$.

51. What is the modulus of $Q(\lambda)$ for $\lambda = 2$?

52. Let $K \subset \mathbb{C}$ be a closed Jordan disk. Then for all $R \gg 0$, $A_R = \Delta(R) - K$ is an annulus. Show that:

$$\lim_{R \to \infty} 2\pi \operatorname{mod}(A_R) - \log R = -\log \operatorname{cap}(K).$$

53. Let $f : A \to B$ be an analytic map between annuli that is also a topological covering map of degree d. Show that $\operatorname{mod}(B) = d \operatorname{mod} A$.

54. Let $A = \{z : 1 < |z| < 2\}$. Show that for any $\epsilon > 0$, there exists a smooth Jordan curve that cuts A into a pair of annuli A_1 and A_2 with $\operatorname{mod}(A_1) + \operatorname{mod}(A_2) < \epsilon$.

55. For each $\lambda > 1$ we have an annulus

$$B(\lambda) = \widehat{\mathbb{C}} - ([0,1] \cup [\lambda, \infty]).$$

(i) Show that the unique closed hyperbolic geodesic in $B(\lambda)$ — its core curve — is a circle.

(ii) Show that $\operatorname{mod} B(\lambda) \cdot \operatorname{mod} Q(\lambda) = 1/2$.

56. (Continuation) Show that $\operatorname{mod} B(\lambda) \sim \log \lambda/(2\pi)$ as $\lambda \to \infty$. Give a proof based on extremal length, or the integral formulas in equation (6.2), or both. You may use the preceding exercise. (Hint: compare this example to equation (6.3).)

57. Let $V \subset \mathbb{Z}^d$ be a finite set, with ∂V defined by equation (7.6), and let $\overline{V} = V \cup \partial V$. Define $P : \mathbb{R}^{\overline{V}} \to \mathbb{R}^{\overline{V}}$ by

$$Pf(x) = \frac{1}{2d} \sum_{|x-y|=1} f(y)$$

if $x \in V$, and $Pf(x) = f(x)$ otherwise. We say f is *harmonic* if $Pf = f$.

Show that for any $f : \overline{V} \to \mathbb{R}$, the limit

$$g(x) = \lim_{n \to \infty} P^n f(x)$$

exists for every $x \in V$, and that g is the unique harmonic function that agrees with f on ∂V. (You may assume the solution to the Dirichlet problem for V, which is outlined in Problem 32.)

(Hint: restrict P to functions that vanish on ∂V, show it is self–adjoint and its eigenvalues are all less than one.)

58. Polya's constant $P = 0.3405\ldots$ is the probability that a random walk in $\mathbb{Z}^3$, starting at the origin, returns to the origin at least once. Prove that P satisfies

$$\frac{1}{1-P} = (2\pi)^{-3} \int_{(S^1)^3} \frac{d\theta}{1 - (1/3)\sum_{i=1}^3 \cos(\theta_i)}.$$

(Hint: using equation (7.5), show that the integral above computes $S = \sum_0^\infty r_3(n)$, and that $S = \sum_0^\infty P^n$.)

59. Adapt the proof of Theorem 7.9 to show that a bounded harmonic function $f(x)$ on $\mathbb{R}^d$ is constant. (Hint: show that, for every $t \in \mathbb{R}^d$, the bounded harmonic $f(x+t) - f(x)$ must achieve its maximum, and hence it is constant. You may use Theorem 3.1.)

60. Let $x(i)$ be a random walk on $\mathbb{Z}$ with $x(0) = 0$. Show that the probability that $x(i) \geq 0$ for $i = 1, 2, \ldots, 2n$ is the same as the probability that $x(i) \neq 0$ for $i = 1, 2, \ldots, 2n$.

61. A plane with $n \geq 2$ seats is boarding n passengers. Passenger p has a reservation for seat p, $p = 1, 2, \ldots, n$, and passenger p prefers seat p, provided $p \geq 2$.

Unfortunately, passenger $p = 1$ insists on choosing a seat at random. If a subsequent passenger finds their reserved seat occupied, they also choose an open seat at random.

Show that the probability that passenger n sits in seat n is $1/2$.

62. Let $V \subset \mathbb{Z}^d$ be a finite set with boundary $B = \partial V$. Let $\tau(x)$ be the least n such that $x(n) \in B$, given that $x(0) = x$; and let $f(x) = E(\tau(x))$.

(i) Show that $Af(x) = -1$ for all $x \in V$.

(ii) Show that, in terms of equation (7.12), we have

$$f(x) = \sum_{y \in V} \sum_0^\infty p_n^B(x, y).$$

Explain why this formula holds, directly from the definition of $p_n^B(x, y)$.

63. (Gambler's ruin revisited.) Consider a random walk $x(n)$ on $\mathbb{Z}$, modeling a game of fair roulette. Fix $N > 0$, and restrict attention to $x = x(0) \in [0, N]$. Let $\tau \geq 0$ be the least index such that $x(\tau) = 0$ or $x(\tau) = N$. Let $\tau' = \tau$ if $x(\tau) = N$, and 0 otherwise. Recall that $s(x) = P_x(x(\tau) = N) = x/N$.

(i) Let $f(x) = E_x(\tau')$. Show that

$$f(x) = s(x) + (f(x+1) + f(x-1))/2,$$

and $f(0) = f(N) = 0$.

(ii) Show that
$$f(x) = x(N^2 - x^2)/(3N).$$

(iii) Let $x = 1000$ and $N = 1050$. We have seen (cf. equation 7.10) that for these values, $E_x(\tau) = 50,000$. Show that the average number of spins it takes complete the game, given that the gambler *wins*, is $E_x(\tau')/s(x)$, and compute its value.

(iv) What is the average number of spins it takes to *lose* the game?

64. Prove that $P(\tau(1) = 2n - 1) = f_{2n}$. Deduce that $P(\tau(1) = n) \sim Cn^{-3/2}$.

65. Let $z(n) = x(n) + iy(n)$ be a random walk on $\mathbb{Z}^2 = \mathbb{Z}[i]$. Let $p(n)$ be the number of times the walk $x(n)$ pauses up to time n; that is, the number of $i \leq n$ such that $x(i) = x(i+1)$.

(i) Prove that $p(n)/n \to 1/2$ with probability one.

(ii) Prove that the probability $q(n)$ that $p(n)/n \leq 1/3$ is very small, e.g. that $q(n) = O(1/n^k)$ for every integer $k > 0$.

66. Let $x(n)$ be a random walk on $\mathbb{Z}$.

(i) Prove that for each $\alpha > 0$, $x(n)/n^{1/2+\alpha} \to 0$ with probability one. (Hint: check that $\int_1^\infty \exp(-t^\alpha)\, dt < \infty$.)

(ii) Prove that $\limsup x(n)/\sqrt{n} = \infty$ almost surely.

67. (i) For $s \geq 0$, evaluate

$$E(s) = \lim_{n\to\infty} \prod_{0<k<s\sqrt{n}} (1 + k/n) \cdot$$

141

(ii) Show that the limit in (i) is uniform on any compact set, $|s| \le M$.

(iii) Sketch a proof of Theorem 8.1, the central limit theorem, using (i) and (ii) and equation (8.2).

68. Let $V \subset \mathbb{Z}^d$ be a finite set. Let $f : V \cup \partial V \to \mathbb{R}$ be a superharmonic function, in the sense that $Af(x) \le 0$ for all $x \in V$. Suppose $\partial V = A \cup B$, $f \ge 0$ on A and $f \ge 1$ on B.

Show that the probability that a random walk $x(n)$ starting at $x \in V$ hits B before it hits A is bounded above by $f(x)$.

69. (i) Let $U \subset \mathbb{C}$ be the sector defined by $\arg(z) \in [\pi/4, \pi/2]$. Show that the function $f(z) = \arg(z)$ is superharmonic on $U \cap \mathbb{Z}[i]$, meaning $Af(z) \le 0$.

(ii) Let $z(n)$ be a random walk on $\mathbb{Z}[i]$ starting at $z(0) = i + N(i+1) \in U$. Show that the probability that $z(n)$ hits the imaginary axis before exiting U is $O(1/N)$.

(iii) Let $z = x + iy$, and let S be the square $[-N, N] \times [-N, N] \subset \mathbb{C}$. Let $z(n)$ be a random walk on $\mathbb{Z}[i]$ starting at $z(0) = i$. Show that the probability that $z(n)$ exits S without crossing the line $x = y$ is $O(1/N)$.

(This variant of Theorem 9.2 can also be used to prove Theorem 9.1 (Smirnov).)

70. Let $f : \Delta \to [-1, 1]$ be a harmonic function. Prove that there exists a universal constant C such that $|df/dx(0)| \le C$. What is the best possible value for C? What function(s) $f(z)$ achieve this value?

71. State and prove a theorem that shows every smooth harmonic function $f(z)$ on a domain in $\mathbb{C}$ is locally the limit of discrete harmonic functions.

72. Prove Lemma 10.4.

73. Let μ_1, μ_2 be two probability measures on $C[0, 1]$, and let $S \subset [0, 1]$ be a countable dense set. For any finite ordered set $E = \{e_1, \ldots, e_n\}$ with $e_i \in S$, we have a map $\pi_E : C[0, 1] \to \mathbb{R}^n$ sending $f(x)$ to $(f(e_1), \ldots, f(e_n))$.

(i) Show that if the pushforwards of μ_1 and μ_2 to $\mathbb{R}^n$ under π_E agree for every possible E, then $\mu_1 = \mu_2$.

(ii) Prove the uniqueness part of Theorem 10.1 for Brownian motion $B(t)$, $t \in [0, 1]$.

74. Let G be an infinite connected graph of degree d with basepoint $x \in V(G)$. (The degree condition means there are d edges incident to each vertex; e.g. the checkerboard graph has degree 4.)

    (i) Show that the number μ_n of simple paths in G, starting at x and of length n, is at most $d(d-1)^{n-1}$.

    (ii) Let $S \subset V(G)$ be a set of vertices chosen independently with probability p. Let S_x be the cluster containing x (this is empty if $x \notin S$). Show that

    $$E(|S_x|) \leq \sum_0^\infty \mu_n p^{n+1}.$$

    (iii) Show that the critical probability for percolation in G satisfies $1/(d-1) \leq p_c(G)$.

    (iv) Let G be the 'bifurcating tree', whose root x has degree two and whose other edges have degree 3. Show that $p_c(G) = 1/2$.

    (Hint: let q_n be the probability that $S \subset V(G)$ includes a path from x to one of the 2^n vertices at distance n from x. Show that $q_0 = p$ and $q_{n+1} = f_p(q_n)$, where $f_p(x) = px(2-x)$. If $\lim q_n > 0$, then $p \geq p_c$.)

References

[Ah1] L. Ahlfors. *Complex Analysis*. McGraw-Hill, 1953.

[Ah2] L. Ahlfors. *Conformal Invariants: Topics in Geometric Function Theory*. McGraw-Hill Book Co., 1973.

[BR] B. Bollobás and O. Riordan. *Percolation*. Cambridge University Press, 2006.

[CFL] R. Courant, K. Friedrichs, and H. Lewy. Über die partiellen Differenzengleichungen der mathematischen Physik. *Math. Ann.* **100** (1928), 32–74.

[DS] H. Duminil-Copin and S. Smirnov. Conformal invariance of lattice models. In *Probability and Statistical Physics in Two and More Dimensions*, volume 15 of *Clay Math. Proc.*, pages 213–276. Amer. Math. Soc., 2012.

[DY] E. B. Dynkin and A. A. Yushkevich. *Markov Processes: Theorems and Problems.* Plenum Press, 1969.

[Fe] W. Feller. *An Introduction to Probability Theory and its Applications.* John Wiley & Sons, Inc., 1968.

[Fer] J. Ferrand. Fonctions préharmoniques et fonctions préholomorphes. *Bull. Sci. Math.* **68** (1944), 152–180.

[LeG] J.-F. Le Gall. *Brownian Motion, Martingales and Stochastic Calculus.* Springer–Verlag, 2013.

[Gol] G. M. Goluzin. *Geometric Theory of Functions of a Complex Variable.* Amer. Math. Soc, 1969.

[Mil] J. Milnor. *Dynamics in One Complex Variable*, volume 160 of *Annals of Math. Studies.* Princeton University Press, 2006.

[MP] P. Mörters and Y. Peres. *Brownian Motion.* Cambridge University Press, 2010.

[Neh] Z. Nehari. *Conformal Mapping.* McGraw-Hill, 1952.

[Rol] D. Rolfsen. *Knots and Links.* Publish or Perish, Inc., 1976.

[Sch] K. Schiefermayr. Zolotarev's conformal mapping and Chebotarev's problem. *Integral Transforms Spec. Funct.* **26** (2015), 118–133.

[Sm] S. Smirnov. Critical percolation in the plane: conformal invariance, Cardy's formula, scaling limits. *C. R. Acad. Sci. Paris Sér. I Math.* **333** (2001), 239–244.

[Str] K. Strebel. On quadratic differentials with closed trajectories and second order poles. *J. Analyse Math.* **19** (1967), 373–382.